

VISIT...

LANZAROTE
Caliente.COM

МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ АВИАЦИОННЫЙ ИНСТИТУТ
(технический университет)

У.Г. ПИРУМОВ

ЧИСЛЕННЫЕ МЕТОДЫ

*Рекомендовано министерством
общего и профессионального образования
Российской Федерации
в качестве учебного пособия
для студентов высших технических
учебных заведений*



Москва
Издательство МАИ
1998

ОГЛАВЛЕНИЕ

Рецензенты

кафедра «Математическая физика» факультета ВМК МГУ,
Н.Н. Калиткин

П 33 **Пирумов У.Г.**
Численные методы Учебное пособие – М Изд-во МАИ, 1998,
– 188 с ил
ISBN 5-7035-2190-4

В книге предлагаются основные численные методы решения задач линейной и нелинейной алгебры, приближения функций, обыкновенных дифференциальных уравнений и уравнений в частных производных, которые автор в течение многих лет читает студентам факультета «Прикладная математика и физика» Московского государственного авиационного института (технического университета)

Книга предназначена для студентов технических вузов, аспирантов и преподавателей, а также инженеров и научных работников, использующих в практической деятельности численные методы

П 1602120000 – 339
094(02) – 98

ISBN 5-7035-2190-4

ББК 16 2 12

© У.Г. Пирумов, 1998

Предисловие	5
Введение	6
Глава 1 Численные методы алгебры	9
1 1 Краткое содержание главы	9
1 2 Некоторые сведения из линейной алгебры. Понятие о нормах	9
1 3 Основные трудности решения систем линейных уравнений. Классификация методов решения	14
1 4 Метод исключения Гаусса	17
1 5 Вычисление определителя и обратной матрицы методом исключения	19
1 6 Метод прогонки для решения систем линейных уравнений с трехдиагональной матрицей	20
1 7 Итерационные методы решения линейных уравнений	23
1 8 Решение нелинейных уравнений. Два этапа отыскания корня	30
1 9 Метод половинного деления (метод дихотомии)	32
1 10 Метод простой итерации нахождения корней нелинейных уравнений	33
1 11 Метод Ньютона (метод линеаризации, метод касательных) и некоторые его модификации	36
1 12 Метод секущих (метод хорд)	39
1 13 Метод парабол (метод квадратичной интерполяции). Метод Мюллера для уточнения корней нелинейных уравнений	40
1 14 Методы нахождения корней систем нелинейных уравнений. Ускорение сходимости по Эйткену	41
1 15 Введение в проблему собственных значений	44
1 16 Итерационный метод вращения для нахождения собственных значений	45
1 17 Метод Данильзского для построения характеристического многочлена матрицы	48
1 18 Метод интерполяции для построения характеристического многочлена	49
Глава 2 Приближение функций. Численное интегрирование и дифференцирование	52
2 1 Краткое содержание главы	52
2 2 Постановка задачи о приближении (аппроксимации) функций	52
2 3 Интерполяция	53
2 4 Интерполяционный многочлен в форме Лагранжа	55
2 5 Интерполяционный многочлен в форме Ньютона	56
2 6 Погрешность и сходимость интерполяции	59
2 7 Интерполяция сплайнами	62
2 8 Приближение методом наименьших квадратов (МНК)	65
2 9 Постановка задачи численного дифференцирования	67
2 10 Дифференцирование интерполяционного многочлена Ньютона	68
2 11 Безразностные формулы численного дифференцирования для равноотстоящих узлов	71
2 12 Применение ряда Тейлора для численного дифференцирования	72
2 13 Постановка задачи численного интегрирования	73
2 14 Формулы численного интегрирования прямоугольников, трапеций, парабол (Симпсона)	73
2 15 Метод Рунге-Ромберга-Ричардсона повышения порядков точности	78

Глава 3. Численные методы решения обыкновенных дифференциальных уравнений

3 1 Краткое содержание главы	80
3 2 Некоторые сведения из теории обыкновенных дифференциальных уравнений (ОДУ)	80
3 3 Понятие о методе конечных разностей. Порядок точности разностной схемы	84
3 4 Метод Эйлера. Метод Эйлера с пересчетом	87
3 5 Метод Рунге-Кутты	89
3 6 Многошаговый метод Адамса	91
3 7 Неявные схемы. Понятие о жестких системах	92
3 8 Постановка краевой задачи для обыкновенного дифференциального уравнения	95
3 9 Метод стрельбы	96
3 10 Конечно-разностный метод решения краевых задач	97

Глава 4. Численные методы решения уравнений в частных производных

4 1 Краткое содержание главы	99
4 2 Некоторые сведения из теории уравнений в частных производных	99
4 3 Основные понятия метода сеток. Задача Дирихле для уравнения Лапласа	106
4 4 Явные и неявные разностные схемы	111
4 5 Аппроксимация, устойчивость, сходимость разностных схем. Основные понятия	113
4 6 Примеры неустойчивых разностных схем	114
4 7 Практические правила исследования устойчивости	118
4 8 Спектральный признак устойчивости	118
4 9 Принцип максимума	119
4 10 Метод гармоник Фурье исследования устойчивости разностных схем	119
4 11 Применение метода гармоник Фурье для исследования устойчивости разностных схем для уравнения переноса	120
4 12 Применение метода гармоник Фурье в задачах исследования устойчивости разностных схем для уравнения теплопроводности	123
4 13 Метод переменных направлений (метод дробных шагов)	125
4 14 Метод установления – математические и физические основы метода	129
4 15 Разностные схемы метода установления	133
4 16 Методы сквозного счета	137
4 17 Метод прямых	141
4 18 Метод характеристик	143
4 19 Метод конечных элементов	147

Литература	154
Великие математики	155

ПРЕДИСЛОВИЕ

В учебном пособии изложены основные численные методы решения прикладных задач. Книга построена по традиционному принципу и содержит четыре главы. Первая глава посвящена численным методам алгебры, вторая – приближением функций, численному дифференцированию и интегрированию, третья – численным методам решения обыкновенных дифференциальных уравнений, четвертая – численным методам решения уравнений в частных производных. Каждая глава сопровождается введением, в котором дается краткое содержание главы, а также приводятся сведения из разделов высшей математики, необходимые для лучшего понимания материала. В некоторых учебных пособиях по численным методам имеются также главы по решению интегральных уравнений и численным методам оптимизации. Автор не включил эти разделы в настоящее пособие, поскольку методы решения интегральных уравнений и численные методы оптимизации частично приводятся в первой главе в разделах, посвященных решению линейных и нелинейных уравнений. Кроме того, численные методы оптимизации зачастую выделены в отдельный курс.

Материал, изложенный в книге, соответствует программам дисциплин по специальности «Прикладная математика» и представлен на основе лекций, читаемых автором в течение последних лет на факультете «Прикладная математика и физика» Московского государственного авиационного института (технического университета). В некоторых разделах пособия помимо традиционных методов рассматриваются численные методы, развитые автором и его коллегами.

Учебное пособие заканчивается главой, в которой приведены сведения о научной деятельности некоторых наиболее известных математиков мира, внесших основополагающий вклад в развитие математической науки.

Автор считает своим долгом выразить глубокую признательность рецензенту члену-корреспонденту РАН, профессору Н. Н. Калиткину, рецензенту профессору А. М. Денисову, а также профессорам В. И. Кирееву, Г. С. Рослякову, В. Ф. Тишкину, В. Ф. Формалеву, доцентам Н. А. Архангельскому, П. В. Бирюкову, В. А. Горяйнову, И. Э. Иванову, С. Б. Карасеву за внимательное прочтение рукописи книги и ценные замечания. Автор выражает благодарность за подготовку рукописи к изданию О. О. Бородину, В. Н. Камзолову, А. Н. Козелько, Т. А. Тихоновой, С. А. Чугункову.

ВВЕДЕНИЕ

В настоящее время численные методы являются мощным математическим средством решения многих научно-технических проблем. Это связано как с невозможностью в большинстве случаев получить точное аналитическое решение, так и со стремительным развитием ЭВМ. Несмотря на существование многочисленных стандартных программ и объектно ориентированных пакетов прикладных программ, для научных и инженерно-технических работников необходимо понимание существа основных численных методов и алгоритмов, поскольку зачастую интерпретация результатов расчетов нетривиальна и требует специальных знаний особенностей применяемых методов. В этой связи важно понимать структуру погрешностей при решении конкретных задач и корректность вычислений.

Структура погрешности

Существуют четыре источника погрешностей, полученные в результате численного решения: математическая и физическая модели, исходные данные, приближенность метода и ошибки округления.

Первые два источника погрешностей приводят к так называемой неустранимой погрешности. Эта погрешность может присутствовать, даже если решение сформулированной задачи найдено точно. Погрешность метода возникает из-за того, что точный оператор и исходные данные, в частности начальные и краевые условия, заменяются по определенным правилам приближенными. Так, производные заменяются их разностными аналогами, интегралы – суммами, функции – специальными многочленами; а при решении многих задач строятся бесконечные итерационные процессы, которые естественным образом прекращаются после конечного числа итераций. Как правило, погрешность метода может быть оценена и поддается контролю. Ниже для некоторых методов такая оценка будет представлена.

Погрешность метода следует выбирать так, чтобы она была не более чем на порядок меньше неустранимой погрешности.

Погрешность округления возникает в связи с тем, что вычисления производятся с конечным числом значащих цифр. Округления производятся по следующему правилу: если в старшем из отбрасываемых разрядов стоит цифра меньше пяти, то содержимое сохраняемых разрядов не изменяется. В противном случае в младший сохраняемый разряд добавляется единица с тем же знаком, что и у самого числа. Очевидно, что погрешность, возникающая при округлении, не превышает младшего оставляемого разряда. Повторное округление проводить не следует, так как оно может привести к увеличению погрешности. Различают абсолютную и относительную погрешность. Пусть a – точное,

вообще говоря, неизвестное числовое значение некоторой величины, а $\bar{a}$ – известное приближенное числовое значение этой величины, тогда число

$$\Delta(\bar{a}) = |a - \bar{a}|$$

называют абсолютной погрешностью числа $\bar{a}$, а величину

$$\delta(\bar{a}) = \frac{\Delta(\bar{a})}{|\bar{a}|}$$

– его относительной погрешностью. Нетрудно показать, что при сложении и вычитании складываются абсолютные погрешности, а при делении и умножении – относительные погрешности. Очевидно, что абсолютная погрешность характеризуется числом верных цифр после запятой, а относительная погрешность – числом верных значащих цифр.

Поскольку на современных ЭВМ число записывается, как правило, минимум с 10 – 12 десятичными знаками, то погрешность единичного округления $\delta = 10^{-10} \div 10^{-12}$ обычно пренебрежимо мала по сравнению с неустранимой погрешностью и погрешностью метода. При решении больших задач производятся миллиарды операций и можно предположить, что ошибки могут заметно накапливаться, однако, поскольку они носят случайный характер, может происходить их взаимная компенсация. Зачастую строятся специальные алгоритмы, в частности итерационные, которые малочувствительны к ошибкам округления.

Корректность

При численном решении основных задач необходимо знать какие-либо входные (исходные) данные – начальные, краевые (граничные) значения искомой функции, коэффициенты и правые части уравнений и т.д. Очевидно, что для исследователя важно знать, существует ли решение поставленной задачи, единственно ли оно и как оно зависит от входных данных.

Говорят, что задача поставлена корректно, если она разрешима при любых допустимых входных данных в случае, когда имеется единственное решение и это решение непрерывно зависит от входных данных, т.е. малому их изменению соответствует малое изменение решения. В этом случае говорят, что задача устойчива.

Задача поставлена некорректно, если её решение неустойчиво относительно входных данных, т.е. их малому изменению могут соответствовать большие изменения решения. Известно, что корректной задачей является задача численного интегрирования, а некорректной – задача дифференцирования.

Классическим примером некорректной задачи является задача Коши для уравнения Лапласа. Эта некорректность исходной задачи проявляется и при её численном решении.

В настоящее время развиты методы решения некорректных задач. К числу их относятся так называемые методы регуляризации, которые сводят решение

исходной задачи к решению близкой к ней вспомогательной с некоторым малым параметром ε , так, что при $\varepsilon \rightarrow 0$ решение вспомогательной задачи должно стремиться к решению исходной задачи. Ниже для некоторых численных методов будут формулироваться условия корректности и устойчивости.

Глава 1. ЧИСЛЕННЫЕ МЕТОДЫ АЛГЕБРЫ

1.1. КРАТКОЕ СОДЕРЖАНИЕ ГЛАВЫ

Методы, приведенные в этой главе, в значительной степени объединяет система понятий и идентичность подходов к решению задач.

Сначала приводятся численные методы решения систем линейных алгебраических уравнений. В линейной алгебре эту задачу называют первой основной задачей. К этой задаче примыкают задачи вычисления определителей и вычисления элементов обратной матрицы. Иногда вычисление определителей и элементов обратной матрицы называют второй и третьей основными задачами линейной алгебры.

Далее излагаются численные методы уточнения корней нелинейных уравнений. В общем случае такие уравнения не имеют аналитических формул для своих корней или эти формулы слишком громоздки и неудобны для практического использования. В частности, в начале прошлого века было доказано, что нелинейные уравнения выше четвертой степени неразрешимы в радикалах, даже если использовать радикалы произвольной степени. Используемые численные методы уточнения корней нелинейных уравнений можно условно разделить на две группы. К первой группе относятся методы, которые назовем универсальными в том смысле, что в принципе они пригодны для отыскания корней уравнений любого вида. Эти методы носят итерационный характер. Подробно рассмотрены метод дихотомии (деление пополам), метод простой итерации, метод Ньютона и две его модификации: метод секущих и метод Ньютона с постоянным значением производной. Кратко дан метод парабол.

Для системы нелинейных уравнений приведем обобщение методов простой итерации, метода итераций Зейделя и метода Ньютона, рассмотренных для одного нелинейного уравнения.

Последняя тема этой главы посвящена задаче отыскания собственных значений и собственных векторов. Эту задачу в линейной алгебре, как правило, называют второй основной задачей. Будут рассмотрены и два метода отыскания всех собственных значений: итерационный метод вращения и метод интерполяции. Дополнительные сведения о методах решения полной проблемы собственных значений содержатся в разделе, где рассматривается метод Данилевского.

1.2. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ЛИНЕЙНОЙ АЛГЕБРЫ. ПОНЯТИЕ О НОРМАХ

Приведем вначале некоторые сведения из линейной алгебры, которые потребуются в дальнейшем. Рассмотрим прямоугольную матрицу

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

Две матрицы $A \equiv [a_{ij}]$ и $B \equiv [b_{ij}]$ размерности $m \times n$ равны друг другу, если $a_{ij} = b_{ij}$ для всех i и j .

Сумма двух матриц A и B размерности $m \times n$ есть матрица размерности $m \times n$:

$$A + B \equiv [a_{ij}] + [b_{ij}] \equiv [a_{ij} + b_{ij}].$$

Произведение матрицы A на скаляр α есть матрица размерности $m \times n$:

$$\alpha A \equiv \alpha [a_{ij}] = [\alpha a_{ij}].$$

Произведение матрицы A размерности $m \times n$ на матрицу B размерности $n \times r$ есть матрица C размерности $m \times r$:

$$C = AB \equiv [a_{ij}][b_{jk}] \equiv [c_{ik}], \text{ где } c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}.$$

Таким образом, элемент c_{ik} матрицы C есть сумма произведений i -й строки матрицы A на соответствующие элементы k -го столбца матрицы B , при этом число столбцов матрицы A должно равняться числу строк матрицы B . Из существования произведения AB вовсе не следует существование произведения BA . Для квадратных матриц ($m = n$) одного порядка существуют AB и BA , но, вообще говоря, $AB \neq BA$.

Определитель (детерминант) квадратной матрицы будем обозначать $\det A$:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Определитель квадратной матрицы с n^2 (действительными или комплексными) числами (элементами a_{ij}) есть сумма $n!$ членов вида $(-1)^r a_{1j_1} a_{2j_2} \dots a_{nj_n}$:

$$\det A = \sum_{j=1}^n a_{ij} A_{ij},$$

где A_{ij} — алгебраическое дополнение элемента a_{ij} . Отметим, что

$$(\det A)^2 \leq \prod_{i=1}^n \prod_{j=1}^n |a_{ij}|^2 \quad (\text{неравенство Адамара}).$$

Важным частным случаем квадратной матрицы является диагональная матрица:

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

При $a_{11} = a_{22} = \dots = a_{nn} = 1$ матрица называется единичной и обозначается через E . Другим частным случаем квадратной матрицы является трехдиагональная матрица:

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \dots & 0 & 0 \\ a_{21} & a_{22} & a_{23} & \dots & 0 & 0 \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & 0 & \dots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

С такими матрицами часто приходится иметь дело при решении дифференциальных уравнений. Матрица называется нижней треугольной, если все ее элементы, расположенные выше главной диагонали, равны нулю:

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 & 0 \\ a_{21} & a_{22} & \dots & 0 & 0 \\ \dots & & & & \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

Аналогично определяется верхняя треугольная матрица. Квадратная матрица называется симметричной, если ее элементы удовлетворяют соотношению $a_{ij} = a_{ji}$, $i, j = 1, 2, \dots, n$. Если в матрице поменять строки со столбцами, то получим транспонированную матрицу:

$$A^T = A' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & & & \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix}.$$

Очевидно, что $x'(x^T)$ обозначает вектор-строку.

Квадратная матрица A симметрична, если $A = A'$.

Матрица называется положительно определенной, если $x'Ax > 0$.

Матрица $\bar{A}$ называется комплексно сопряженной с матрицей A , если ее элементы суть комплексно сопряженные с элементами матрицы A . Матрица $A^* = \bar{A}'$ называется сопряженной с матрицей A . Если A вещественная, то $A^* = A'$.

Матрица A^{-1} называется обратной матрице A , если $AA^{-1} = E$.

Необходимым и достаточным условием существования обратной матрицы A^{-1} является условие $\det A \neq 0$. Говорят, что в этом случае матрица A несобственная, или невырожденная.

Минором порядка k матрицы A называется определитель k -го порядка, составленный из элементов, которые находятся на пересечении k строк и k столбцов матрицы A . Рангом матрицы A называется такое число r , что все миноры порядка $r+1$ и выше равны нулю.

Характеристическим уравнением матрицы A называется уравнение

$$\det \begin{bmatrix} a_{11} - \lambda & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} - \lambda & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} - \lambda \end{bmatrix} = 0.$$

Корни λ , этого уравнения называются собственными числами матрицы A . Левая часть уравнения $\det(A - \lambda E)$ называется характеристическим полиномом. Собственным вектором матрицы A называется отличный от нуля вектор, удовлетворяющий условию

$$Ax = \lambda x.$$

Две квадратные матрицы A и B называются подобными, если $B = S^{-1}AS$, где S – невырожденная матрица и $A = SBS^{-1}$. Подобные матрицы имеют одинаковые характеристические многочлены и, следовательно, одинаковые собственные значения и определители, так как $\det A = \prod_{i=1}^n \lambda_i$. Собственные векторы

матрицы A связаны с собственными векторами матрицы B соотношением $x = By$. Матрица A невырождена, если все λ_i отличны от 0. Все собственные числа симметричной матрицы действительны.

Действительная матрица называется ортогональной, если ее транспонированная матрица совпадает с обратной, т. е. $A'(A^T) = A^{-1}$ или $AA' = A'A = E$. Все собственные значения ортогональной матрицы по модулю равны единице. Строки (столбцы) ортогональной матрицы попарно ортогональны, суммы квадратов элементов каждой строки (столбца) ортогональной матрицы равны единице, определитель ортогональной матрицы равен ± 1 ; если матрица A ортого-

нальна, то и матрица A^{-1} тоже ортогональна.

Всякая симметричная действительная матрица может быть приведена к диагональному виду подобным преобразованием

$$A = UAU^{-1},$$

где U – ортогональная матрица; A – диагональная матрица. Из свойств ортогональной матрицы следует, что

$$A = U^{-1}AU.$$

В дальнейшем зачастую рассматриваются множества, элементами которых являются числа, векторы, матрицы, функции. Сами множества обычно являются линейными нормированными пространствами, ибо в них определены операции сложения элементов и их умножения на число и введена норма каждого элемента $\|x\|$. Нормой называется вещественное неотрицательное число, удовлетворяющее следующим условиям:

1. $\|x\| > 0$ при $x \neq 0$, $\|x\| = 0$ при $x = 0$.
2. $\|ax\| = |a| \|x\|$.
3. $\|x + y\| \leq \|x\| + \|y\|$.

Рассмотрим нормы в некоторых множествах.

В множестве действительных чисел $\|x\| = |x|$.

В множестве функций $x(t)$, определенных и непрерывных при $0 \leq t \leq 1$ (пространство C), определена чебышевская норма $\|x\|_c = \max |x(t)|$.

В конечномерном пространстве, элементами которого являются группы из n чисел $x = \{x_1, x_2, \dots, x_n\}$ (их можно считать координатами векторов), норма определена так:

$$\|x\|_p = \left[\frac{1}{n} \sum_{i=1}^n |x_i|^p \right]^{\frac{1}{p}}.$$

Между разными нормами существует соотношение $\|x\|_1 \leq \|x\|_2 \leq \|x\|_c \leq \sqrt{n} \|x\|_2 \leq n \|x\|_1$, где $\|x\|_c = \sup |x_i|$ по аналогии с пространством C .

Поэтому из сходимости в одной из норм следует сходимость в остальных.

В пространстве квадратных матриц порядка n наиболее употребительны следующие нормы:

$$\|A\|_c = \max_i \left[\sum_{j=1}^n |a_{ij}| \right], \quad \|A\|_1 = \max_j \left[\sum_{i=1}^n |a_{ij}| \right], \quad \|A\|_M = n \max_{i,j} |a_{ij}|, \quad \|A\|_E = \left[\sum_{i,j=1}^n |a_{ij}|^2 \right]^{1/2}$$

Норма $\|A\|_c$ – максимальное число среди сумм модулей элементов строк матрицы, норма $\|A\|_l$ – максимальное число среди сумм модулей элементов столбцов матрицы. Эти нормы не имеют специального названия, третья норма $\|A\|_M$ называется максимальной, а четвертая $\|A\|_E$ – сферической или евклидовой. Отметим, что $\|A+B\| \leq \|A\| + \|B\|$; $\|AB\| \leq \|A\| \cdot \|B\|$.

Интересна связь между нормами матриц и векторов, на которые матрицы действуют. Норма матрицы называется согласованной с нормой вектора, если $\|Ax\| \leq \|A\| \cdot \|x\|$. Сферическая норма согласована с $\|x\|_2$, а максимальная согласована со всеми рассмотренными выше нормами.

Введем понятие предела векторов и матриц. Рассмотрим последовательность векторов $x^{(1)}, x^{(2)}, \dots$ с компонентами $x_1^{(1)}, \dots, x_n^{(1)}, x_1^{(2)}, \dots, x_n^{(2)}, \dots$. Если существуют пределы

$$\lim_{k \rightarrow \infty} x_i^{(k)}, \quad i = 1, 2, \dots, n,$$

то говорят, что вектор с компонентами $x_1, x_2, \dots, x_n$ является пределом последовательности $x^{(1)}, x^{(2)}, \dots$. Аналогично определяется предел последовательности матриц. Необходимым и достаточным условием сходимости векторов $x^{(k)}$ к x является условие $\|x^{(k)} - x\| \rightarrow 0$, при этом $\|x^{(k)}\| \rightarrow \|x\|$. Аналогичное утверждение справедливо для матриц.

1.3. ОСНОВНЫЕ ТРУДНОСТИ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ. КЛАССИФИКАЦИЯ МЕТОДОВ РЕШЕНИЯ

Требуется найти решение системы линейных уравнений

$$Ax = b, \quad (1.1)$$

где $A = (a_{ij})$ – квадратная матрица коэффициентов при неизвестных; $x = (x_j)$ – вектор-столбец неизвестных; $b = (b_j)$ – вектор-столбец правых частей системы. С точки зрения классической теории линейных алгебраических систем их решение не вызывает затруднений. По правилу Крамера система n линейных уравнений с n неизвестными имеет единственное решение, если определитель системы отличен от нуля ($\det A \neq 0$) и значение каждого из неизвестных вычисляется как отношение двух определителей порядка n , т.е.

$$x_j = \det A_j / \det A, \quad j = 1, \dots, n, \quad (1.2)$$

где $\det A_j$ – определитель матрицы, получаемой заменой j -го столбца матрицы A столбцом правых частей. При непосредственном вычислении определителей

как алгебраической суммы $n!$ произведений элементов для отыскания решения системы линейных уравнений по правилу Крамера требуется приблизительно $n \cdot n!$ арифметических операций типа умножения. Тогда ЭВМ с быстродействием около 5 млн. опер./с потратит на вычисление решения системы линейных уравнений порядка $n = 20$ около шести миллионов лет. Мы увидим, что использование же метода исключения Гаусса позволяет уменьшить время, необходимое для решения задачи (1.1), до величины менее одной секунды.

Другое важное обстоятельство, связанное с решением систем линейных алгебраических уравнений, состоит в следующем. С точки зрения теории линейных систем различаются два случая: определитель матрицы системы не равен нулю ($\det A \neq 0$), т.е. система уравнений является невырожденной, и определитель матрицы системы равен нулю ($\det A = 0$) – вырожденная система. Во втором случае система либо не имеет решения (при $b \neq 0$), либо имеет неединственное решение (при $b = 0$). С точки зрения практических вычислений существуют «почти невырожденные системы» – системы, у которых определитель близок к нулю, но отличен от нуля ($\det A \approx 0$). Небольшие изменения коэффициентов матрицы системы или правых частей системы в «почти невырожденных системах» могут привести к большим погрешностям решения.

Все эти случаи хорошо иллюстрируются на примере решения системы двух линейных уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1, \\ a_{21}x_1 + a_{22}x_2 = b_2. \end{cases}$$

На рис. 1.1. каждому уравнению соответствует прямая на плоскости x_1x_2 , а точка пересечения этих прямых есть решение системы (1.1).

Если $\det A = 0$, то наклоны прямых равны и они либо параллельны, либо совпадают. При $\det A \approx 0$ небольшие погрешности в коэффициентах и правых частях могут привести к большим погрешностям в решении, т.е. к положению точки пересечения.

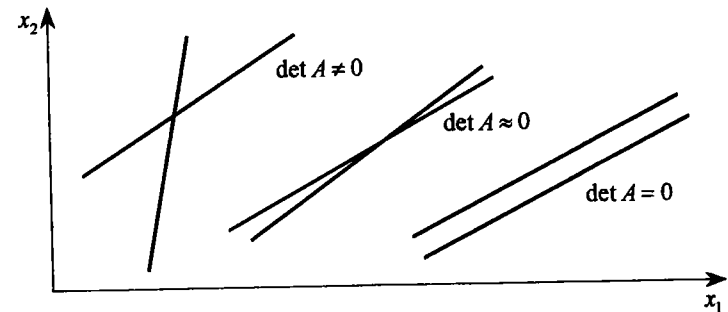


Рис. 1.1

Системы такого типа, в которых есть малые погрешности в коэффициентах системы или в правых частях (эти погрешности могут быть, в частности, результатом округлений при вычислениях или записи чисел в память ЭВМ), называются плохо обусловленными.

Плохо обусловленная система геометрически соответствует почти параллельным прямым. В теоретических исследованиях обусловленность характеризуется числом обусловленности $\chi = \|A\| \cdot \|A^{-1}\|$, при любой норме $\chi \geq 1$. Чем больше это число, тем хуже обусловленность системы; так, при $\chi \approx 10^3 + 10^4$ система уже плохо обусловлена. На практике, однако, ограничиваются проверкой условия $\det A \approx 0$.

Системы линейных уравнений можно было бы рассматривать как частный случай нелинейных уравнений. Однако относительная простота линейных систем обусловила появление специальных высокоэффективных методов их решения. По этой же причине введение и пояснение основных понятий могут быть выполнены относительно просто. Вместе с тем линейные системы представляют большой самостоятельный интерес по двум причинам. Во-первых, к системам линейных уравнений приводят многие практические задачи. Во-вторых, важным и неиссякаемым источником систем линейных алгебраических уравнений являются практически все разделы вычислительной математики: нахождение решения системы линейных алгебраических уравнений необходимо в задачах уточнения корней систем нелинейных уравнений, аппроксимации функций, отыскания собственных значений и собственных векторов матрицы, решения обыкновенных дифференциальных уравнений и уравнений в частных производных и т.д. Методы решения линейных уравнений и метод прогонки используются и для некоторых классов интегральных уравнений и интегродифференциальных уравнений.

Будут рассмотрены наиболее употребительные прямые и итерационные методы. Прямые методы дают решение задачи за конечное (точно определяемое для каждого метода) число операций. Здесь намеренно не употребляется термин «точные методы», так как из-за ошибок округления при вычислениях с конечным числом знаков решение всегда получается с погрешностями. Причины этого и способы уменьшения влияния ошибок округления будут обсуждены ниже. Из прямых методов рассмотрены метод Гаусса, метод Гаусса с выбором главного элемента для системы линейных алгебраических уравнений общего вида, метод прогонки для систем линейных уравнений с трехдиагональной матрицей.

Итерационные методы дают решение как предел бесконечной последовательности приближенных решений, в которых каждое последующее более точное приближение находится по уже найденному предыдущему (или предыдущим решениям). Из итерационных методов решения систем линейных алгебраических уравнений рассмотрены метод простой итерации и метод Зейделя.

1.4. МЕТОД ИСКЛЮЧЕНИЯ ГАУССА

Систему уравнений (1.1) представим в виде

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n, \end{cases} \quad (1.3)$$

$$\text{или } \sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n.$$

Известно большое число схем метода исключения, приспособленных для ручного или машинного счета матриц общего или специального вида.

Метод Гаусса можно интерпретировать как метод, в котором первоначально матрица приводится к верхней треугольной форме (прямой ход), а далее – к единичной (обратный ход). Очевидно, что если матрица единичная, то $x_i = b_i$.

Пусть матрица системы (1.3) – верхняя треугольная, поэтому $a_{ij} = 0$ при $i > j$, т.е. все элементы ниже главной диагонали равны нулю. Тогда из последнего уравнения сразу определяем x_n . Подставляя его в предпоследнее уравнение, находим x_{n-1} и т.д.

Общие формулы имеют вид

$$\begin{aligned} x_k &= \frac{b_k}{a_{kk}}, \quad k = n, \\ x_k &= \frac{1}{a_{kk}} \left[b_k - \sum_{l=k+1}^n a_{kl}x_l \right], \quad k = n-1, n-2, \dots, 1, \\ \text{при } k > l \quad a_{kl} &= 0. \end{aligned} \quad (1.4)$$

Приведем матрицу системы (1.3) к верхней треугольной. Вычтем из второго уравнения системы (1.3) первое, умноженное на такое число, чтобы коэффициент при x_1 обратился в нуль.

То же сделаем со всеми остальными уравнениями. В результате все коэффициенты первого столбца, лежащие ниже главной диагонали, обратятся в нуль. Затем, используя второе уравнение, обратим в нуль соответствующие коэффициенты второго столбца. Последовательно продолжая этот процесс, приведем матрицу системы к верхней треугольной форме.

Запишем общие формулы метода Гаусса. Пусть проведено исключение коэффициентов из $k-1$ столбца. Тогда останутся уравнения с ненулевыми элементами ниже главной диагонали:

$$\sum_{j=k}^n a_{ij}^{(k)} x_j = b_i, \quad k \leq i \leq n.$$

Умножим k -ю строку на число $c_{mk} = a_{mk}^{(k)} / a_{kk}^{(k)}$, $m > k$ и вычтем из m -й строки. Первый ненулевой элемент этой строки обратится в нуль, а остальные изменятся по формулам

$$\begin{aligned} a_{ml}^{(k+1)} &= a_{ml}^{(k)} - c_{mk} a_{kl}^{(k)}, \\ b_m^{(k+1)} &= b_m^{(k)} - c_{mk} b_k^{(k)}, \quad k < m. \end{aligned}$$

Проведя вычисления по этим формулам при всех указанных индексах, обратим в нуль элементы k -го столбца, лежащие ниже главной диагонали. Аналогичная процедура приводит матрицу системы к верхней треугольной форме, при этом весь процесс приведения называется прямым ходом метода Гаусса. Вычисление неизвестных по формулам (1.4) называют обратным ходом метода.

Обратный ход можно совершить иначе, если обратить в нуль и все коэффициенты, лежащие выше главной диагонали. Например, элементы n -го столбца обращаются в нуль, если $a_{nn}^{(n)}$ умножить на $-a_{kn}^{(n)} / a_{nn}^{(n)}$ и сложить с соответствующей строкой. Аналогично обращаются в нуль и все остальные столбцы. Если, кроме того, разделить затем каждое уравнение на соответствующий элемент, стоящий на главной диагонали, то матрица системы становится единичной, а неизвестные $x_i = b_i^{(2n)}$, где $b_i^{(2n)}$ – коэффициенты правой части i -го уравнения, после указанных преобразований.

На некотором шаге прямого хода может оказаться, что коэффициент $a_{ii}^{(k)} \neq 0$, но мал по сравнению с остальными элементами матрицы системы и, в частности, мал по сравнению с элементами первого столбца. Деление коэффициентов системы на малую величину может привести к значительным ошибкам округления.

Для уменьшения ошибок округления поступают следующим образом. Среди элементов первого столбца $a_{ik}^{(k)}$ каждой промежуточной матрицы выбирают наибольший по модулю (главный) элемент и путем перестановки i -й строки и строки, содержащей главный элемент, делают главный элемент ведущим. Такая модификация метода исключения Гаусса называется методом Гаусса с выбором главного элемента. Случай появления нулевых элементов обходится при этом сам собой.

Для реализации метода требуется примерно $n^3/3$ операций типа умножения и $n^3/3$ операций типа сложения. Полезно помнить, что оценка числа опе-

раций определяется в основном операциями, затрачиваемыми при выполнении прямого хода метода Гаусса. Обратный ход метода Гаусса требует примерно n^2 операций. Следовательно, если требуется решить несколько систем линейных алгебраических уравнений $Ax = b$ с одной и той же матрицей и различными правыми частями, то общее число операций при решении S систем будет оцениваться величиной $\frac{2}{3}n^3 + Sn^2$. В этом случае целесообразно реализовать алгоритм метода Гаусса в виде двух подпрограмм: первая подпрограмма должна реализовывать прямой ход алгоритма и получать на выходе верхнюю треугольную матрицу, а вторая подпрограмма должна, используя полученную матрицу, вычислять решение системы для произвольной правой части.

1.5. ВЫЧИСЛЕНИЕ ОПРЕДЕЛИТЕЛЯ И ОБРАТНОЙ МАТРИЦЫ МЕТОДОМ ИСКЛЮЧЕНИЯ

Определитель треугольной матрицы равен произведению диагональных элементов. В результате выполнения прямого хода метода исключения система линейных уравнений приводится к верхней треугольной матрице. Следовательно, определитель матрицы системы может быть вычислен как произведение ведущих элементов:

$$\det A = (-1)^k \prod_{i=1}^n a_{ii}^{(i)},$$

где k – количество перестановок строк при использовании метода исключения с выбором главного элемента. Вычисление определителя требует примерно $(2/3)n^3$ операций.

Для вычисления обратной матрицы вспомним, что $AA^{-1} = E$, где A^{-1} – обратная матрица, а E – единичная матрица. Обозначим элементы обратной матрицы через α . Тогда соотношение $AA^{-1} = E$ можно записать в виде

$$\sum_{k=1}^n a_{ik} \alpha_{ki} = \delta_{il}, \quad 1 \leq i \leq n, \quad (1.5)$$

где δ_{il} – символ Кронекера, $\begin{cases} \delta_{il} = 1, & i = l, \\ \delta_{il} = 0, & i \neq l. \end{cases}$

Видно, что если рассматривать l -й столбец обратной матрицы как вектор, то он является решением системы (1.5) с матрицей A со специальной правой частью, в которой на l -м месте стоит единица, а остальные – нули.

Решение этой системы для каждого l дает элементы l -го столбца обратной матрицы. Таким образом, для обращения матрицы необходимо решить n систем линейных уравнений n -го порядка с одинаковой матрицей A и различными пра-

выми частями. Так как приведение матрицы A к треугольной выполняется только один раз, то для обращения матрицы с использованием метода исключения требуется примерно $2n^3$ операций.

1.6. МЕТОД ПРОГОНКИ ДЛЯ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ С ТРЕХДИАГОНАЛЬНОЙ МАТРИЦЕЙ

Системы линейных уравнений с трехдиагональной матрицей коэффициентов при неизвестных являются наиболее важным и распространенным случаем систем специального вида. В таких системах отличны от нуля только элементы, лежащие на главной диагонали и на нижней и верхней диагоналях, прилегающих к ней. К системам с трехдиагональными матрицами приводят, например, задачи о сплайн-интерполяции, о решении разностными методами обыкновенных дифференциальных уравнений и уравнений в частных производных.

Метод прогонки принадлежит к числу прямых методов решения систем линейных уравнений и используется в тех случаях, когда многие коэффициенты матрицы равны нулю. Это обстоятельство учтено при реализации метода прогонки, в котором исключаются преобразования с нулевыми элементами. В методе прогонки применительно к системе линейных уравнений, имеющих трехдиагональную матрицу, можно выделить следующие этапы:

1. Приведение трехдиагональной матрицы к верхней треугольной (прямой ход). В случае трехдиагональной матрицы это означает приведение к двухдиагональной, т.е. приведение исходной системы к системе, содержащей по два неизвестных в каждом уравнении, кроме последнего, в котором только одно неизвестное.

2. Запись обратного хода в виде $x_i = P_{i+1}x_{i+1} + Q_{i+1}$, так как преобразованная матрица – двухдиагональная.

3. Вывод рекуррентного соотношения для P_{i+1} и Q_{i+1} через P_i и Q_i и получение соотношения для P_2 и Q_2 ($P_1 = Q_1 = 0$).

4. Осуществление обратного хода метода прогонки и определение всех неизвестных.

Рассматриваемый ниже метод прогонки представляет собой модификацию метода исключения Гаусса, использующую специальный регулярный вид матрицы системы. Запишем систему линейных алгебраических уравнений с трехдиагональной матрицей в виде

$$\begin{aligned} a_i x_{i-1} - b_i x_i + c_i x_{i+1} &= d_i, \quad i = 1, 2, \dots, n, \\ a_1 &= c_n = 0. \end{aligned} \quad (1.6)$$

Это так называемый канонический вид системы линейных уравнений метода прогонки.

При этом матрица системы (1.6) имеет вид

$$A = \begin{bmatrix} -b_1 & c_1 & 0 & \dots & 0 & 0 \\ a_2 & -b_2 & c_2 & \dots & 0 & 0 \\ \dots & a_3 & -b_3 & c_3 & \dots & 0 \\ 0 & 0 & 0 & \dots & -b_{n-1} & c_{n-1} \\ 0 & 0 & 0 & \dots & a_n & -b_n \end{bmatrix}.$$

Прямой ход метода прогонки сводится к исключению неизвестного x_{i-1} в каждом уравнении системы. Получаемая в результате прямого хода система содержит в каждом уравнении только два неизвестных x_i и x_{i+1} , и матрица её – верхняя треугольная с двумя диагоналями. Запишем i -ю строку преобразованной двухдиагональной матрицы в виде

$$x_i = P_{i+1}x_{i+1} + Q_{i+1}. \quad (1.7)$$

Если система (1.6) приведена к виду (1.7), то обратный ход метода Гаусса очевиден. Однако использование общих алгоритмов прямого и обратного хода не целесообразно. Построим эффективную вычислительную схему, которая и составляет суть метода прогонки. Для этого, уменьшив в (1.7) индекс на единицу, запишем

$$x_{i-1} = P_i x_i + Q_i.$$

Подставляя это соотношение в (1.6), имеем

$$a_i(P_i x_i + Q_i) - b_i x_i + c_i x_{i+1} = d_i,$$

откуда нетрудно получить

$$x_i = [b_i - a_i P_i]^{-1} c_i x_{i+1} + [b_i - a_i P_i]^{-1} [a_i Q_i - d_i].$$

Сравнивая это соотношение с (1.7), можем записать рекуррентные соотношения

$$P_{i+1} = c_i [b_i - a_i P_i]^{-1}, \quad Q_{i+1} = [a_i Q_i - d_i] [b_i - a_i P_i]^{-1} \quad (1.8)$$

для вычисления так называемых прогоночных коэффициентов.

Подчеркнем, что последующие значения прогоночных коэффициентов P_{i+1} , Q_{i+1} вычисляются только по известным коэффициентам системы (1.6) и известным предыдущим значениям прогоночных коэффициентов P_i , Q_i .

Для начала прямого хода метода прогонки необходимо задать начальные (стартовые) значения прогоночных коэффициентов, например P_1 , Q_1 . Отметим,

что, вообще говоря, значения коэффициентов P_1, Q_1 в рассмотренной схеме вычислений не требуются, так как значения коэффициентов P_2, Q_2 вычисляются только через коэффициенты первого уравнения системы: при $i=1$ формула (1.6) принимает вид $-b_1x_1 + c_1x_2 = d_1$, при сравнении её с формулой (1.7) при $i=1$ имеем $P_2 = c_1/b_1$; $Q_2 = -d_1/b_1$, а значение x_1 в обратном ходе вычисляется по соотношению $x_1 = P_2x_2 + Q_2$. Использование P_1, Q_1 в качестве начальных значений целесообразно по двум причинам: сохраняется однородность вычислительного алгоритма для всех $i = 2, 3, \dots, n$; упрощается обсуждение и доказательство условия корректности и устойчивости метода прогонки. Из того факта, что P_2, Q_2 не зависят от P_1, Q_1 (так как в соотношениях (1.8) при $i=1$ коэффициенты P_1, Q_1 умножаются на $a_1 = 0$), следует, что можно задать любые значения для прогоночных коэффициентов P_1, Q_1 . Из последующего будет ясно, почему удобно положить $P_1 = Q_1 = 0$. Для начала обратного хода метода прогонки необходимо для вычисления $x_i = P_{i+1}x_{i+1} + Q_{i+1}$ задать значение x_{n+1} . Так как $c_n = 0$, то из первого соотношения (1.8) вытекает, что $P_{n+1} = 0$ и, следовательно, можно задать любое значение для x_{n+1} . Обычно полагают $x_{n+1} = 0$, и тогда $x_n = Q_{n+1}$.

Метод прогонки устойчив, если $|P_i| \leq 1$. Метод прогонки корректен, если $b_i - a_iP_i \neq 0$.

Отметим, что условие устойчивости означает не возрастание ошибок округления, а их подавление. Пусть $\tilde{x}_i, \tilde{x}_{i+1}$ – вычисленные с погрешностями значения решения. Пусть при этом P_{i+1}, Q_{i+1} вычисляются точно. Тогда из (1.7)

$$\tilde{x}_i - x_i = P_{i+1}(\tilde{x}_{i+1} - x_{i+1})$$

или

$$\varepsilon_i = \tilde{x}_i - x_i = P_{i+1}\varepsilon_{i+1}.$$

Из этой формулы при $|P_{i+1}| \leq 1$ следует, что в этом случае погрешность не возрастает.

Достаточным условием корректности метода прогонки и устойчивости его к погрешностям является условие преобладания диагональных коэффициентов:

$$|b_i| \geq |a_i| + |c_i|, \quad i = 1, 2, \dots, n.$$

В самом деле, если хотя бы для одного i выполняется строгое неравенство $|P_i| < 1$, то можно записать цепочку неравенств:

$$|P_{i+1}| = |c_i| / |b_i - a_iP_i| \leq |c_i| / [|b_i| - |a_i||P_i|] \leq |c_i| / [|a_i| + |c_i| - |a_i||P_i|] < 1.$$

Выше было показано, что можно положить $P_1 = 0$ и тогда, во-первых, для

всех i выполняется $|P_i| < 1$, что обеспечивает затухание погрешности, и, во-вторых, для всех i выполняется условие $|b_i - a_iP_i| > |c_i| \geq 0$ и, таким образом, не возникает деления на нуль. Так как условие $|b_i| \geq |a_i| + |c_i|$ является только достаточным, то невыполнение его не означает нарушения корректности и устойчивости.

Для реализации метода требуется примерно 8*n* операций, из которых 3*n* операций типа умножения и 5*n* операций типа сложения. При численном решении дифференциальных уравнений используются различные варианты метода прогонки: метод встречных прогонок, потоковая прогонка, матричная прогонка для систем векторных уравнений.

$$\text{Отметим, что } \det A = \prod_{i=1}^n [a_iP_i - b_i].$$

Метод прогонки обобщается на случай, когда в системе (1.6) a_i, b_i, c_i – квадратные матрицы размерности $m \times m$, а x_i, d_i – векторы размерностью m . Тогда соотношения (1.7) и (1.8) метода прогонки сохраняются, в которых все действия совершаются уже над матрицами, а не над скалярами; а в (1.8) $[b_i - a_iP_i]^{-1}$ является соответствующей обратной матрицей.

Условие устойчивости матричной прогонки: $|\det P_i| < 1$, а корректности и устойчивости:

$$\det(B_i^{-1}C_i) + \det(B_i^{-1}A_i) < 1.$$

1.7. ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ ЛИНЕЙНЫХ УРАВНЕНИЙ

В итерационных методах предполагается осуществление трех следующих этапов: построение итерационного процесса для вычисления последовательных приближений, сходящегося к точному решению (т.е. построение последовательности векторов $x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(k)}$, сходящихся к точному решению x^*); определение критерия сходимости построенного процесса, позволяющего определить момент достижения требуемой точности; исследование скорости сходимости и оптимизация итерационного процесса с целью уменьшения числа операций, необходимых для достижения требуемой точности.

Итерационные методы обладают свойством, позволяющим получить решение с наперед заданной точностью, если доказана сходимость метода. Принципиально точного решения итерационные методы не дают, поскольку оно достигается как предел последовательности векторов. Прямой метод принципиально дает точное решение, но из-за ошибок округления, имеющих место в любых ЭВМ, оно не может быть достигнуто и априори даже трудно оценить, насколько это решение отличается от точного. В связи с отмеченным иногда итерацион-

ные методы позволяют получить решение с большей точностью, чем прямые.

Ниже рассмотрим несколько итерационных методов решения линейных уравнений.

Метод простой итерации. В методе простой итерации система (1.1) линейных алгебраических уравнений $Ax = b$ приводится к эквивалентной системе вида

$$x = \alpha x + \beta. \quad (1.9)$$

Решение системы (1.9) и, следовательно, решение исходной системы (1.1) ищется как предел при $k \rightarrow \infty$ последовательности векторов

$$x^{(k+1)} = \alpha x^{(k)} + \beta, \quad k = 0, 1, 2, \dots, \quad (1.10)$$

где $x^{(0)}$ – начальное приближение для вектора решения.

Достаточное условие сходимости метода простой итерации определяется следующей теоремой.

Теорема 1. Если какая-либо норма матрицы α , согласованная с рассматриваемой нормой вектора x , меньше единицы ($\|\alpha\| < 1$), то последовательность $x^{(k)}$ в методе простой итерации сходится к точному решению x^* системы (1.9) со скоростью не меньше скорости геометрической прогрессии со знаменателем $q \leq \|\alpha\|$ при любом начальном приближении $x^{(0)}$.

Доказательство. Для доказательства теоремы введем погрешность $\varepsilon^{(k)} = x^* - x^{(k)}$. Вычитая из равенства $x^* = \alpha x^* + \beta$ равенство (1.10), получаем $\varepsilon^{(k+1)} = \alpha \varepsilon^{(k)}$. Переходя к нормам, имеем

$$\|\varepsilon^{(k)}\| = \|\alpha^k \varepsilon^{(0)}\| \leq \|\alpha^k\| \cdot \|\varepsilon^{(0)}\| \leq \|\alpha\|^k \cdot \|\varepsilon^{(0)}\|.$$

Отметим, что неравенство $\|\alpha^k \varepsilon^{(0)}\| \leq \|\alpha^k\| \cdot \|\varepsilon^{(0)}\|$, входящее в предыдущую формулу, есть условие согласованности нормы матрицы и вектора. Если $\|\alpha\| < 1$, то при любом векторе начальной погрешности (или иначе – при любом начальном векторе $x^{(0)}$) норма погрешности $\|\varepsilon^{(k)}\|$ стремится к нулю не медленнее геометрической прогрессии со знаменателем $q \leq \|\alpha\|$. Заметим, что верхний индекс k в скобках означает номер итерации, а без скобок – как обычно, степень.

Если в качестве нормы матрицы выбрать норму $\|\alpha\|_c = \max_i \sum_{j=1}^n |\alpha_{ij}|$ или

$\|\alpha\|_1 = \max_j \sum_{i=1}^n |\alpha_{ij}|$, то для решения вопроса о сходимости метода простой итерации можно воспользоваться следствием из теоремы: метод простой итерации

сходится, если для матрицы α выполняется одно из следующих условий:

$$\begin{aligned} \sum_{j=1}^n |\alpha_{ij}| &< 1, \quad i = 1, 2, \dots, n, \\ \sum_{i=1}^n |\alpha_{ij}| &< 1, \quad j = 1, 2, \dots, n. \end{aligned} \quad (1.11)$$

Простейшим и распространенным способом приведения системы $Ax = b$ к виду (1.9), удобному для итераций, является выделение диагональных элементов, при этом каждое i -е уравнение разрешается относительно i -го неизвестного:

$$a_{ii}x_i = \left[b_i - \sum_{j=1, j \neq i}^n a_{ij}x_j \right], \quad i = 1, 2, \dots, n \quad (1.12)$$

и метод простой итерации запишется в виде

$$a_{ii}x_i^{(k+1)} = \left[b_i - \sum_{j=1, j \neq i}^n a_{ij}x_j^{(k)} \right].$$

Матрица α при этом имеет вид

$$\begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \dots & -\frac{a_{2n}}{a_{22}} \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & -\frac{a_{nn-1}}{a_{nn}} & 0 \end{bmatrix}.$$

Элемент этой матрицы можно записать как $\alpha_{ij} = -\frac{a_{ij}}{a_{ii}}[1 - \delta_{ij}]$, где δ_{ij} –

символ Кронекера.

В этом случае достаточное условие сходимости метода простой итерации может быть сформулировано как условие преобладания диагональных элементов матрицы A , что следует из (1.11) и записи матрицы α , т.е.

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

Еще раз подчеркнем, что рассмотренные формы условия сходимости метода итерации являются лишь достаточными. Их выполнение гарантирует сходимость метода, но их невыполнение в общем случае не означает, что метод

простой итерации расходится. Необходимое и достаточное условие сходимости метода простой итерации есть условие, что целая часть $|\lambda_{\max}| < 1$ (где $\lambda_{\max}$ – максимальное по модулю собственное значение матрицы A); оно редко используется в практике вычислений.

Перейдем к вопросу об оценке погрешности решения. Представляют интерес два соотношения оценки погрешности решения $\varepsilon^{(k)} = x^* - x^{(k)}$: первое связывает норму погрешности с нормой разности двух последовательных приближений $\|x^{(k)} - x^{(k-1)}\|$ и может быть использовано для оценки погрешности только в процессе вычислений; второе связывает норму погрешности с нормами вектора начального приближения $\|x^{(0)}\|$ и вектора свободного члена $\|\beta\|$ в системе (1.9). Необходимые соотношения даются следующими двумя теоремами.

Теорема 2. Если какая-либо норма матрицы α , согласованная с рассматриваемой нормой вектора x , меньше единицы ($\|\alpha\| < 1$), то имеет место следующая оценка погрешности:

$$\|\varepsilon^{(k)}\| \leq \frac{\|\alpha\|}{1 - \|\alpha\|} \|x^{(k)} - x^{(k-1)}\|. \quad (1.13)$$

Доказательство. Вычтем из равенства $x^* = \alpha x^* + \beta$ равенство (1.10). Получим

$$x^* - x^{(k)} = \alpha [x^* - x^{(k-1)}].$$

Вычитая справа и слева $x^{(k-1)}$, преобразуем это соотношение к виду

$$x^* - x^{(k-1)} = [x^{(k)} - x^{(k-1)}] + \alpha [x^* - x^{(k-1)}].$$

Перейдя к нормам, запишем

$$\|\varepsilon^{(k-1)}\| \leq \|x^{(k)} - x^{(k-1)}\| + \|\alpha\| \cdot \|\varepsilon^{(k-1)}\|,$$

или

$$[1 - \|\alpha\|] \cdot \|\varepsilon^{(k-1)}\| \leq \|x^{(k)} - x^{(k-1)}\|.$$

Так как по условию теоремы $[1 - \|\alpha\|] > 0$, то можем записать

$$\|\varepsilon^{(k-1)}\| \leq \frac{\|\alpha\|}{1 - \|\alpha\|} \|x^{(k)} - x^{(k-1)}\|.$$

Используя соотношение $\varepsilon^{(k)} = \alpha \varepsilon^{(k-1)}$, из которого следует, что $\|\varepsilon^{(k)}\| \leq \|\alpha\| \cdot \|\varepsilon^{(k-1)}\|$, окончательно получим

$$\|\varepsilon^{(k)}\| \leq \frac{\|\alpha\|}{1 - \|\alpha\|} \|x^{(k)} - x^{(k-1)}\|.$$

Теорема 3. Если какая-либо норма матрицы α , согласованная с рассматриваемой нормой вектора x , меньше единицы ($\|\alpha\| < 1$), то имеет место следующая оценка погрешности:

$$\|\varepsilon^{(k)}\| \leq \|\alpha\|^k \cdot \|x^{(0)}\| + \frac{\|\alpha\|^k}{1 - \|\alpha\|} \|\beta\|.$$

Сделаем два замечания. Во-первых, соотношение (1.13) может быть записано в виде

$$\|\varepsilon^{(k)}\| \leq \frac{\|\alpha\|^k}{1 - \|\alpha\|} \|x^{(k)} - x^{(0)}\|,$$

позволяющем получить оценку погрешности по результатам двух первых итераций. Во-вторых, при использовании метода итераций в качестве оценки погрешности вычислений иногда рекомендуется использовать норму разности двух последовательных приближений. Из соотношений для погрешности следует, что в общем случае это неверно. Если $\|\alpha\|$ близка к единице, то коэффициент при $\|x^{(k)} - x^{(k-1)}\|$ может быть достаточно большим.

Погрешности последовательных итераций связаны соотношением

$$\|\varepsilon^{(k+1)}\| = \|\alpha\| \cdot \|\varepsilon^{(k)}\|,$$

т.е. погрешность изменяется на шаге линейно. Говорят, что метод имеет линейную сходимость или первый порядок сходимости. Вместе с тем количество итераций, необходимое для достижения требуемой точности, зависит от значения $\|\alpha\|$ и начального приближения $x^{(0)}$.

Итак, на примере метода простой итерации продемонстрированы три этапа итерационных методов: построение последовательности векторов, порождаемых формулой (1.10); определение условия сходимости по теореме 1 и оценка скорости сходимости с помощью теорем 2 и 3.

Метод Зейделя. В методе простой итерации не используется кажущаяся очевидной возможность улучшения сходимости итерационного процесса – немедленное введение в расчет вновь вычисленных компонент вектора $x^{(k)}$. Эта

возможность используется в итерационном методе Зейделя. Итерационный процесс для системы (1.9) выполняется при этом по соотношению

$$x_i^{(k+1)} = \sum_{j=1}^{i-1} \alpha_{ij} x_j^{(k+1)} + \sum_{j=i+1}^n \alpha_{ij} x_j^{(k)} + \beta_i, \quad i = 1, 2, \dots, n, \quad (1.14)$$

или для системы (1.1)

$$a_{ii} x_i^{(k+1)} = b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)}.$$

Не вдаваясь в подробности, отметим, что метод итераций Зейделя часто действительно приводит к более быстрой сходимости, чем метод простой итерации. Однако возможны случаи, когда метод итераций Зейделя сходится медленнее метода простой итерации, и даже возможны случаи, когда метод простой итерации сходится, а метод итераций Зейделя расходится.

Отметим, что метод Зейделя сходится, если матрица A положительно определена и симметрична.

Покажем, что метод итераций Зейделя эквивалентен некоторому методу простой итерации со специальным образом построенной матрицей α и вектором β в соотношении (1.10). Для этого запишем систему (1.14) в виде $x^{(k+1)} = Hx^{(k+1)} + Fx^{(k)} + \beta$, где F – верхняя треугольная матрица из коэффициентов матрицы α , а $H = \alpha - F$. Перепишем систему в виде $(E - H)x^{(k+1)} = Fx^{(k)} + \beta$, где E – единичная матрица. Матрица $(E - H)$ – нижняя треугольная матрица с диагональными элементами, равными единице. Следовательно, определитель этой матрицы отличен от нуля (равен единице) и она имеет обратную матрицу $(E - H)^{-1}$. Тогда можем записать $x^{(k+1)} = (E - H)^{-1} Fx^{(k)} + (E - H)^{-1} \beta$. Сопоставляя это соотношение с (1.10), можем заключить, что, действительно, метод итераций Зейделя эквивалентен методу простой итерации в том смысле, что для установления условия и критерия сходимости метода итераций Зейделя можно воспользоваться теоремами, установленными для метода простой итерации, если положить $\alpha = (E - H)^{-1} F$. Итерационный процесс для системы (1.12) записывают и в более общей форме, а именно:

$$a_{ii} x_i^{(k+1)} = \omega b_i + (1 - \omega) a_{ii} x_i^{(k)} - \omega \left[\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right],$$

вводя в итерационный процесс для $x_i^{(k+1)}$ значение $x_i^{(k)}$, что не имеет места в методе простой итерации и методе Зейделя.

Итерационный процесс при $\omega > 1$ называют методом верхней релаксации, при $\omega = 1$ (метод Зейделя) – методом полной релаксации и при $\omega < 1$ – методом нижней релаксации.

Для некоторого простого случая при специально выбранном ω можно дать оценку для числа итераций N , необходимого для достижения заданной точности ε .

$$N(\varepsilon) = 0.2n^2 \ln(1/\varepsilon); \quad 0.1n^2 \ln(1/\varepsilon); \quad 0.64n^2 \ln(1/\varepsilon)$$

для метода простой итерации, метода Зейделя и метода верхней релаксации соответственно. Очевидно, что число итераций тем больше, чем больше порядок системы; при этом имеет место квадратичная зависимость для первых двух методов и линейная зависимость для метода верхней релаксации. Очевидно также, что число итераций тем больше, чем меньше ε .

Представленные оценки не являются универсальными, и хотя в обсуждаемом случае самым медленным является метод простой итерации, а самым быстрым – метод верхней релаксации, это соотношение в других случаях может изменяться.

Можно дать наиболее общую запись итерационного процесса для системы (1.1). Имеем

$$B_k \frac{x^{(k+1)} - x^{(k)}}{\tau_k} + Ax^{(k)} = b, \quad (1.15)$$

где B_k – некоторая матрица, выбираемая для обеспечения быстрой сходимости метода, во многом определяется конкретной системой (1.1) и искусством вычислителя. Говорят, что итерационный процесс стационарный, если B_k и τ_k не зависят от k . Из (1.15) очевидно, что если итерационный процесс сходится, т.е. $|x^{(k+1)} - x^{(k)}| \rightarrow 0$, то он сходится к решению системы (1.1).

Пусть матрицы D и L определены как

$$D = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}, \quad L = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & 0 \end{bmatrix}.$$

Тогда метод простой итерации есть частный случай процесса (1.15) при $B = D$ и $\tau = 1$, метод Зейделя – частный случай при $B = D + L$, а метод релаксации – частный случай при $B = D + \omega L$, $\tau = \omega$.

Матрицу B_k следует выбирать как можно ближе к матрице A , но так, чтобы обращение этой матрицы было более простой задачей (треугольная, трехдиагональная и т.д.).

1.8. РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ. ДВА ЭТАПА ОТЫСКАНИЯ КОРНЯ

Пусть имеется нелинейное уравнение

$$f(x) = 0. \quad (1.16)$$

Требуется найти корни этого уравнения, т.е. те значения x , которые обращают уравнение (1.16) в тождество. В процессе приближенного отыскания корней уравнения (1.16) обычно выделяют два этапа: отделение корня и уточнение корня.

Под отделением корня понимается определение промежутка, содержащего один и только один корень уравнения. Одна из точек этого промежутка принимается за начальное приближение корня. В зависимости от метода, который предполагается использовать для уточнения корня, требуется определение тех или иных свойств отделенного корня и поведения функции на отрезке отделения. Например, при использовании простейшего метода уточнения корня – метода дихотомии, необходимо и достаточно установить лишь непрерывность функции на отрезке отделения. При использовании других методов может потребоваться выяснить, является ли корень действительным, какова кратность корня, установить непрерывность и монотонность функции и ее некоторых низших производных.

В общем случае уравнения (1.16) этап отделения корня не может быть алгоритмизирован. Для отдельных классов уравнений (наиболее известным из которых является класс алгебраических уравнений) разработаны специальные приемы отделения корней, существенно облегчающие данное отделение и позволяющие ставить вопрос об автоматизации этого процесса. Некоторые из этих приемов будут приведены при рассмотрении методов решений алгебраических уравнений. Нередко отделение корней нелинейных уравнений выполняется «вручную» с использованием всей возможной информации о функции $f(x)$. В ряде случаев приближенное значение корня может быть определено из физических соображений, если речь идет о решении нелинейного уравнения, связанного с конкретной прикладной задачей. Успешно применяется графический метод отделения действительных корней, обладающий большой наглядностью и позволяющий относительно просто устанавливать возможность наличия кратных корней. При графическом отделении корней бывает полезным представить уравнение (1.16) в эквивалентном виде $f_1(x) = f_2(x)$ и искать точки пересечения функций $f_1(x)$ и $f_2(x)$. Например, для уравнения $f(x) \equiv x^3 - x + 1$ вместо построения графика $y = f(x)$ (рис. 1.2) удобнее и проще построить графики функций $y = f_1(x) = x^3$ и $y = f_2(x) = x - 1$.

В ряде случаев может быть полезным расширение упомянутой выше теоремы: если производная $f'(x)$ существует и сохраняет постоянный знак, то уравнение $f(x) = 0$ имеет на $[a, b]$ единственный корень.

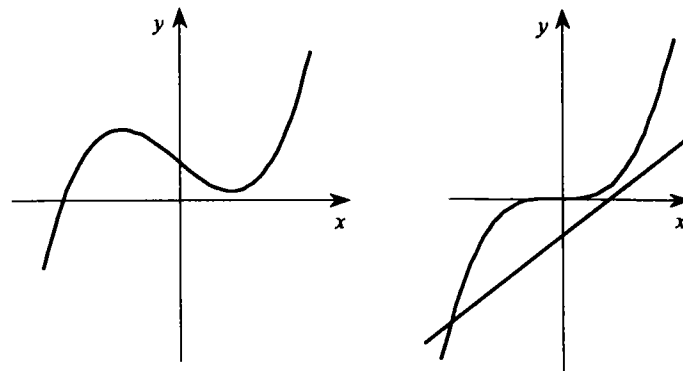


Рис. 1.2

В случае, если на концах интервала функция имеет одинаковые знаки, то на этом интервале корни либо отсутствуют, либо их четное число.

Для корней многочлена n -й степени $P_n(x) = \sum_{i=0}^n a_i x^i$ известно, что интервал, в котором расположены все корни, в том числе и комплексные, определяется по формуле

$$|x_p| \leq 1 + \frac{1}{|a_n|} \max [|a_0|, |a_1|, \dots, |a_{n-1}|].$$

Кроме того, имеет место правило знаков Декарта, согласно которому разность между числом перемен знака в последовательности $a_0, a_1, \dots, a_n$ и числом положительных корней есть положительное число или нуль (в случае действительных корней). Это правило распространяется и на отрицательные корни при замене x на $-x$. Правило Декарта позволяет также оценить число действительных корней на интервале $[a, b]$. Для этого правило знаков нужно применить к уравнению $P(y) = 0$, где

$$x = \frac{a + by}{1 + y}.$$

Полезно также использовать теорему Гюа.

Теорема Гюа. Если для последовательности коэффициентов $a_0, a_1, \dots, a_n$ квадрат каждого крайнего коэффициента больше произведения соседних с ним коэффициентов, т.е. $a_k^2 > a_{k-1} a_{k+1}$, $k = 1, 2, \dots, n-1$, то все корни действительны.

Теорема. Если для каких-либо k выполнено неравенство $a_k^2 < a_{k-1}a_{k+1}$, то многочлен имеет по крайней мере пару комплексных корней.

На втором этапе уточнения при нахождении корня используют два типа методов: прямые и итерационные. В прямых методах корень уравнения принципиально находится за конечное, заранее известное число операций. Прямыми методами удается решить некоторые простейшие алгебраические и тригонометрические уравнения.

В итерационных методах корень x^* определяется как предел некоторой последовательности $x^{(0)}, x^{(1)}, \dots, x^{(k)}$ и решение принципиально не может быть достигнуто за конечное, заранее известное число операций.

Основные методы решения нелинейных уравнений и систем являются итерационными, и к их числу принадлежат метод дихотомии (половинного деления), метод простой итерации, метод Ньютона (метод касательных), метод секущих, метод парабол (метод Мюллера), метод Зейделя. Ниже эти методы будут рассмотрены.

Важной характеристикой итерационных методов является скорость сходимости процесса. Говорят, что метод имеет n -й порядок сходимости, если $|x^{(k+1)} - x^*| = C|x^{(k)} - x^*|^n$, где C – константа, не зависящая от n . При $n=1$ имеем сходимость первого порядка, или линейную сходимость, а при $n=2$ – второго порядка, или квадратичную. Говорят, что метод является одношаговым, если для построения итерационной последовательности нужно вычислить функцию в одной точке, двушаговой – в двух и т.п.

Сравнение различных методов следует проводить по числу операций при реализации одной итерации и по скорости сходимости.

Изложенные ниже методы решения нелинейных уравнений и систем широко используются в численных методах оптимизации.

1.9. МЕТОД ПОЛОВИННОГО ДЕЛЕНИЯ (МЕТОД ДИХОТОМИИ)

Пусть действительный корень уравнения $f(x)=0$ отделен и функция $f(x)$ непрерывна на интервале $[a, b]$ отделения корня. Построим процесс сужения интервала $[a, b]$ такой, чтобы искомый корень всегда находился внутри суженного интервала. Очевидно, что в этом случае погрешность приближенного значения корня не превышает $|b^{(k)} - a^{(k)}|$, где $b^{(k)}, a^{(k)}$ – граничные точки интервала на k -й итерации. Найдем середину отрезка $x^{(1)} = (a+b)/2$ и вычислим $f((a+b)/2)$. Составим произведения $f(a) \cdot f((a+b)/2)$ и $f(b) \cdot f((a+b)/2)$. Из двух половин отрезков выберем тот, где произведение отрицательно, и обозначим новые границы $a^{(1)}, b^{(1)}$. Затем новый отрезок делим пополам, вновь составляем аналогичные произведения и выбираем тот из отрезков, где произведение отрицательно.

Погрешность метода определяется достаточно очевидным соотношением (которое, впрочем, может быть строго доказано)

$$\varepsilon^{(k)} = |x^* - x^{(k)}| \leq |b^{(k)} - a^{(k)}| \leq |b - a|/2^k,$$

которое указывает на скорость сходимости метода: с увеличением k погрешность стремится к нулю не медленнее геометрической прогрессии со знаменателем $q = 1/2$. Дихотомия проста и надежна, всегда сходится, хотя и медленно, устойчива к ошибкам округления. Метод дихотомии, однако, не обобщается на системы уравнений.

1.10. МЕТОД ПРОСТОЙ ИТЕРАЦИИ НАХОЖДЕНИЯ КОРНЕЙ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

При использовании метода простой итерации для уточнения корня уравнение $f(x)=0$ заменяется эквивалентным уравнением

$$x' = \varphi(x), \quad (1.17)$$

т.е. таким, что из $f(x^*)=0$ следует $x^* = \varphi(x^*)$ и наоборот. Привести уравнение (1.16) к (1.17) можно многими способами, например, положив $\varphi(x) = x + \psi(x)f(x)$, где $\psi(x)$ – непрерывная произвольная знакопостоянная функция.

Геометрически уравнение (1.17) на интервале отделения корня представляется в виде двух пересекающихся линий $y = x$ и $y = \varphi(x)$ (рис. 1.3). Полагая, что известно начальное приближение $x^{(0)}$ для значения корня x^* , построим итерационный процесс

$$x^{(k+1)} = \varphi(x^{(k)}), \quad k = 0, 1, 2, \dots, \quad (1.18)$$

изображенный на рис. 1.3 ломаной линией со стрелочками, указывающими направление движения. Видим, что для представленного на рис. 1.3 случая взаимного расположения линий $y = x$ и $y = \varphi(x)$ неограниченное повторение вычислений по соотношению (1.18) позволяет сколь угодно близко подойти к точному значению корня x^* .

Исследуем сходимость метода. Если $\varphi(x)$ имеет непрерывную производную, то из теоремы Лагранжа о конечном приращении:

$$x^{(k+1)} - x^* = \varphi(x^{(k)}) - \varphi(x^*) = [x^{(k)} - x^*] \varphi'(\xi) \quad (1.19)$$

следует, что точка ξ лежит между точками $x^{(k)}$ и x^* . Поэтому, если всюду $|\varphi'(x)| \leq q < 1$, то отрезки $|x^{(k)} - x^*|$ убывают не медленнее геометрической прогрессии со знаменателем $q < 1$. Действительно, из (1.19), которое можно рас-

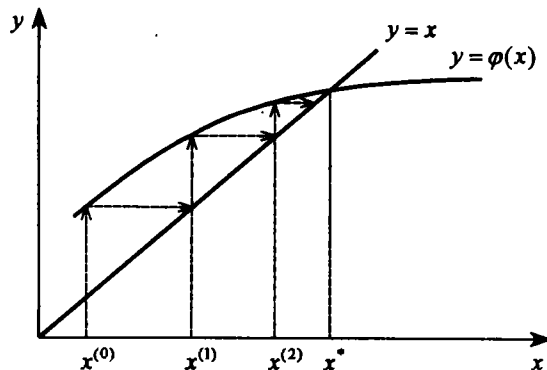


Рис. 1.3

сма́тривать как рекуррентное соотношение, следует, что $|x^{(k)} - x^*| = q^k |x^{(0)} - x^*|$ и последовательность $x^{(k)}$ сходится при любом нулевом приближении.

Итак, условие

$$|\varphi'(x)| < 1 \quad (1.20)$$

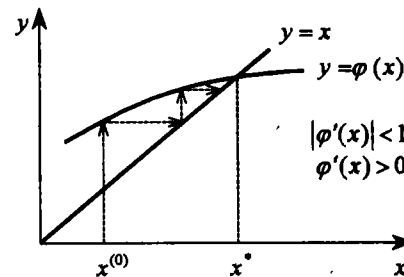
является достаточным условием сходимости итераций. Если $|\varphi'(x)| > 1$, то итерации могут не сходиться. Если $|\varphi'(x)| < 1$, но вдали от корня $|\varphi'(x)| > 1$, то итерации сходятся, если начальное приближение выбрано достаточно близко к корню. При произвольном начальном приближении сходимости может не быть. Таким образом, в методе простой итерации важен выбор начального приближения. Из соотношения (1.19) следует, что если на интервале отделения корня выполняется условие

$$|\varphi'(x)| \leq q < 1,$$

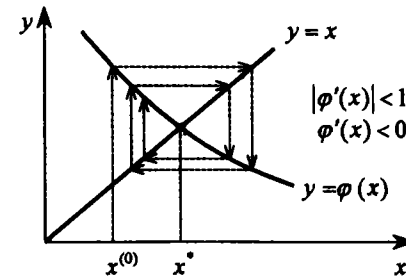
то погрешность на каждой итерации уменьшается для любого начального приближения не медленнее членов геометрической прогрессии со знаменателем q .

Четыре случая взаимного расположения линий $y = x$ и $y = \varphi(x)$ вблизи корня и соответствующие им итерационные процессы показаны на рис. 1.4.

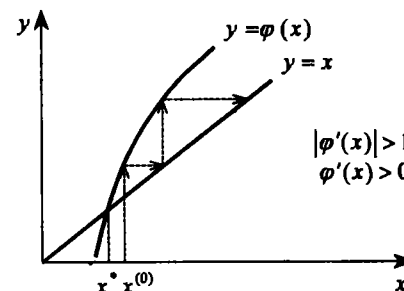
Рис. 1.4, а и 1.4, б соответствуют случаю $|\varphi'(x)| < 1$ – процесс итераций сходится. При этом в первом случае (рис. 1.4, а) $\varphi'(x) > 0$ и сходимость носит односторонний характер, а во втором (рис. 1.4, б) $\varphi'(x) < 0$ и сходимость носит двусторонний характер. Рис. 1.4, в и 1.4, г соответствуют случаю $|\varphi'(x)| > 1$ – процесс итерации расходится, при этом имеет место односторонняя и двусторонняя расходимость.



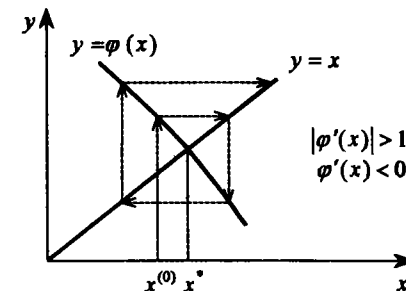
а



б



в



г

Рис. 1.4

Подчеркнем, что приведенное условие (1.20) сходимости метода итераций является лишь достаточным. При этом все приближения должны попадать в отрезок отделения корня. Его выполнение гарантирует сходимость процесса (1.18), но невыполнение условия (1.20) в общем случае не означает, что итерационный процесс будет расходящимся. Например, для случая, показанного на рис. 1.5, условие (1.20) на интервале $[x^*, x^{(0)}]$ не выполняется, но метод итераций сходится.

Используя соотношения (1.19) и (1.20), можем записать

$$|x^* - x^{(k+1)}| = \varepsilon^{(k+1)} \leq q \varepsilon^{(k)} = q |x^* - x^{(k)}|,$$

где $q = \max |\varphi'(x)|$, $x \in [a, b]$. Из последнего соотношения видно, что скорость сходимости метода итерации зависит от величины q : чем меньше q , тем быстрее сходится метод.

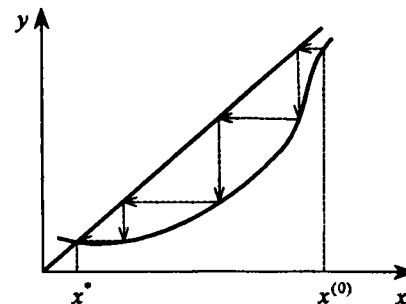


Рис. 1.5

Исходное уравнение $f(x) = 0$ может быть преобразовано к виду $x = \varphi(x)$ многими способами, и, очевидно, для метода итерации целесообразно брать то уравнение $x = \varphi(x)$, для которого q имеет наименьшее значение.

Для пояснения рассмотрим классический пример вычисления квадратного корня. Исходное уравнение $f(x) = x^2 - a = 0$ ($a \geq 0$) преобразуем к виду $x = \varphi(x)$ тремя способами, приведенными в табл. 1.1 в первом столбце.

Таблица 1.1

$\varphi(x)$	$\varphi'(x)$	Поведение $ \varphi'(x) $	Сходимость метода
$\frac{a}{x}$	$-\frac{a}{x^2}$	$ \varphi'(x) \rightarrow 1$ при $x \rightarrow \pm\sqrt{a}$	Не сходится
$x^2 + x - a$	$2x + 1$	$ \varphi'(x) < 1$ при $x \in (-1, 0)$ $ \varphi'(x) \geq 1$ при $x \notin (-1, 0)$	Сходится в ограниченном интервале к отрицательному значению корня
$\frac{(x + a/x)}{2}$	$\frac{(1 - a/x^2)}{2}$	$ \varphi'(x) \rightarrow 0$ при $x \rightarrow \pm\sqrt{a}$	Сходится, и очень быстро

Анализ поведения $|\varphi'(x)|$ вблизи корня (третий столбец таблицы) показывает, что при удачном выборе представления $x = |\varphi(x)|$ можно обеспечить высокую скорость сходимости итерационного процесса без ограничения диапазона параметра a . Третье уравнение $x = (x + a/x)/2$ используется для вычисления квадратного корня на ЭВМ. Таким образом, в методе простой итерации важен выбор вида функций $\varphi(x)$. Отметим, что метод простой итерации обобщается на случай систем линейных уравнений.

1.11. МЕТОД НЬЮТОНА (МЕТОД ЛИНЕАРИЗАЦИИ, МЕТОД КАСАТЕЛЬНЫХ) И НЕКОТОРЫЕ ЕГО МОДИФИКАЦИИ

Вновь рассмотрим уравнение (1.16). Полагая, что погрешность $\varepsilon^{(k)} = x^* - x^{(k)}$ имеет малую величину, а функция $f(x)$ – непрерывную вторую производную, разложим $f(x^*)$ в ряд Тейлора:

$$f(x^*) = f(x^{(k)} + \varepsilon^{(k)}) = f(x^{(k)}) + \varepsilon^{(k)} f'(x^{(k)}) + \frac{(\varepsilon^{(k)})^2}{2} f''(\xi) + \dots,$$

где $\xi \in [x^{(k)}, x^*]$. Учитывая, что $f(x^*) = 0$ и оставляя только линейную часть разложения в ряд (отсюда и название метода – метод линеаризации), можем записать приближенное, линейное относительно погрешности, уравнение

$$f(x^{(k)}) + \varepsilon^{(k)} f'(x^{(k)}) = 0,$$

откуда для погрешности имеем

$$\bar{\varepsilon}^{(k)} = -f(x^{(k)})/f'(x^{(k)}). \quad (1.21)$$

Так как использована лишь линейная часть разложения в ряд, то при подстановке (1.21) в соотношение $x^* = x^{(k)} + \varepsilon^{(k)}$, следующее из соотношения для погрешности, получим вместо x^* лишь приближенное уточненное значение корня, которое обозначим $x^{(k+1)}$. Тогда можем записать основное соотношение метода Ньютона в виде

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}. \quad (1.22)$$

Это соотношение позволяет построить последовательность приближений $x^{(k)}$, $k = 0, 1, 2, \dots$ к точному значению корня по заданному начальному приближению $x^{(0)}$.

Геометрически процесс (1.22) означает замену на каждой итерации кривой $y = f(x)$, касательной к ней в точке $(x^{(k)}, f(x^{(k)}))$, и определение значения $x^{(k+1)}$ как координаты точки пересечения касательной и оси абсцисс (рис. 1.6). С рассмотренной интерпретацией соотношения (1.22) связано название метода – метод касательных.

Достаточное условие сходимости метода Ньютона получим из соответствующего условия для метода простой итерации. Сопоставляя соотношения (1.17) и (1.22), можно видеть, что метод Ньютона представляет собой специальный случай метода простой итерации, в котором

$$\varphi(x) = x - f(x)/f'(x).$$

Используя условие сходимости метода итераций $|\varphi'(x)| < 1$ и выражение $\varphi'(x) = [f(x)f''(x)]/[f'(x)]^2$, нетрудно получить достаточное условие сходимости метода Ньютона в форме

$$|f(x)f''(x)| < [f'(x)]^2. \quad (1.23)$$

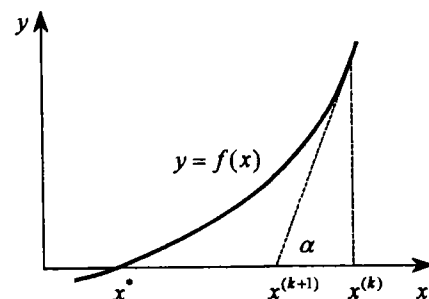


Рис. 1.6

Таким образом, $\varphi'(x^*) = 0$, так как $f(x^*) = 0$.

Так как $f(x^*) = 0$, то $\varphi'(x^*) = 0$ и итерации по соотношению (1.22) сходятся к точному значению корня при произвольном начальном приближении, но вдали от корня сходимость может быть немонотонной.

Для оценки скорости сходимости метода Ньютона запишем соотношение

$$x^{(k+1)} - x^* = \varphi(x^{(k)}) - \varphi(x^*).$$

Далее разложим $\varphi(x^{(k)})$ в ряд Тейлора:

$$\varphi(x^{(k)} - x^* + x^*) = \varphi(x^*) + \varphi'(x^*)(x^{(k)} - x^*) + \frac{1}{2}\varphi''(x^*)(x^{(k)} - x^*)^2 + \dots$$

Подставляя это разложение в предыдущую формулу и учитывая, что $\varphi'(x) = 0$, получаем

$$\varepsilon^{(k+1)} \cong (1/2)\varphi''(x^*)[\varepsilon^{(k)}]^2.$$

Из этого соотношения следует, что метод Ньютона имеет вблизи корня второй порядок сходимости: на каждой итерации ошибка меняется пропорционально квадрату ошибки на предыдущей итерации. Нетрудно видеть, что метод Ньютона является одношаговым. Достоинства метода Ньютона состоят в его квадратичной сходимости, возможности обобщения на случай систем уравнений, а также в том, что он является одношаговым. Однако метод Ньютона расходится в тех областях, где $f'(x) \cong 0$. Кроме того, если функция $f(x)$ задана таблично, то вычисление $f'(x)$ затруднено.

Указанная трудность устраняется в методе секущих (метод хорд).

Существует вариант метода Ньютона с постоянным значением производной; значение производной вычисляется только в начальной точке $x = x^{(0)}$ (рис. 1.7), и далее для всех итераций значения производных $f'(x)$ полагаются постоянными, равными $f'(x^{(0)})$.

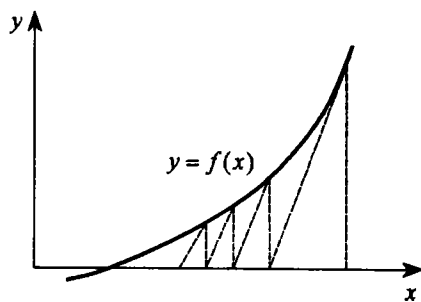


Рис. 1.7

Геометрически замена производной на каждой итерации постоянным значением означает, что наклон прямой для корня уравнения, считается постоянным. Часто эта модификация метода Ньютона рекомендуется как позволяющая уменьшить объем вычислений

за счет отказа от вычисления производной на каждой итерации. Однако следует учитывать, что метод Ньютона с постоянным значением производной имеет лишь первый порядок сходимости – вблизи корня соотношение для погрешности имеет вид

$$\varepsilon^{(k+1)} \approx \varepsilon^{(k)}[1 - f'(x^{(k)})/f'(x^{(0)})]$$

и, следовательно, сходится медленнее, чем метод Ньютона. Это, в конечном счете, нередко приводит к увеличению общего объема вычислений.

1.12. МЕТОД СЕКУЩИХ (МЕТОД ХОРД)

В методе секущих приближенное значение производной $f'(x)$ определяется по двум последовательным приближениям $[x^{(k-1)}, f(x^{(k-1)})]$ и $[x^{(k)}, f(x^{(k)})]$, по соотношению

$$f'(x^{(k)}) \approx \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}, \quad (1.24)$$

что приводит к замене касательной в точке $[x^{(k)}, f(x^{(k)})]$ секущей, проведенной через две точки кривой $y = f(x)$ (рис. 1.8); или, что то же самое, – к аппроксимации функции $f(x)$ на этом интервале линейной функцией.

Условия сходимости метода секущих аналогичны условиям сходимости метода Ньютона. Порядок сходимости метода секущих определяется соотношениями

$$\varepsilon^{(k+1)} \approx \alpha \varepsilon^{(k)} \varepsilon^{(k-1)} \approx \alpha^{1/p} [\varepsilon^{(k)}]^p,$$

где $\alpha = (1/2)f''(\xi)/f'(x^{(k)})$,

$$p = \sqrt{5}/2 \approx 1.618.$$

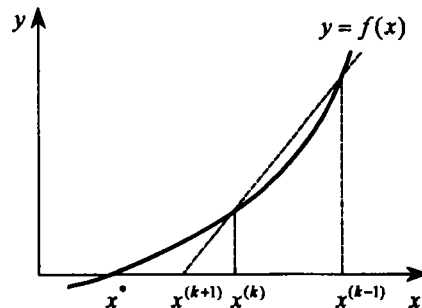


Рис. 1.8

К особенностям метода следует отнести следующее: в методе не требуется непосредственное вычисление производной $f'(x)$ на каждой итерации, что может привести к существенному уменьшению объема вычислений; метод является двухшаговым, и, в частности, на первой итерации вычислений необходимо знать два начальных значения $x^{(0)}$ и $x^{(1)}$; сходимость метода может быть немонотонной даже в малой окрестности корня; в знаменателе формулы для вычисления $x^{(k+1)}$ стоит разность двух величин: $[f(x^{(k)}) - f(x^{(k-1)})]$, которые имеют вблизи корня малые и близкие значения, что может привести к заметным погрешностям вычислений, особенно для кратных корней.

1.13. МЕТОД ПАРАБОЛ (МЕТОД КВАДРАТИЧНОЙ ИНТЕРПОЛЯЦИИ). МЕТОД МЮЛЛЕРА ДЛЯ УТОЧНЕНИЯ КОРНЕЙ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотренный выше метод секущих можно интерпретировать как метод, в котором на каждой итерации исходная функция аппроксимируется линейной функцией (секущей), построенной по двум точкам, принадлежащим $f(x)$. Развивая далее идеи аппроксимации, можно для построения итерационных формул использовать информацию о функции в нескольких точках, предшествующих точке $x^{(k)}$. В методе парабол по трем последовательным приближениям

$$[x^{(k-2)}, f(x^{(k-2)})], [x^{(k-1)}, f(x^{(k-1)})], [x^{(k)}, f(x^{(k)})]$$

строится многочлен второй степени (парабола), приближающий исходную функцию. За новое приближение берется обычно ближайший к $x^{(k)}$ корень соответствующего квадратного уравнения. Геометрическая интерпретация метода парабол дана на рис. 1.9.

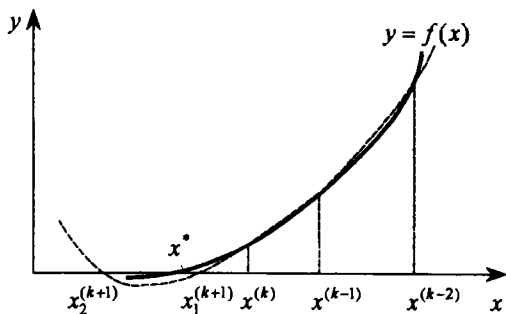


Рис. 1.9

погрешность метода определяется соотношением

$$\varepsilon^{(k+1)} \approx \varepsilon^{(k)} \varepsilon^{(k-1)} \varepsilon^{(k-2)} \approx [\varepsilon^{(k)}]^p, \text{ где } p = 1.839.$$

Это означает, что, несмотря на привлечение дополнительной информации о функции, метод парабол имеет порядок сходимости, лишь немного превышающий порядок сходимости метода секущих. Вместе с тем возникают задачи решения квадратного уравнения, выбора одного из двух корней многочлена и, самое важное, определения области гарантированной сходимости метода. Если три приближения для построения многочлена выбраны далеко от корня и содержат погрешности, то возможно самое неожиданное поведение решения.

Отметим, что метод парабол успешно применяется для отыскания корней многочленов, в том числе комплексных; при этом метод обладает тем замечательным свойством, что начальное приближение может быть действительным. Метод парабол является трехшаговым методом.

1.14. МЕТОДЫ НАХОЖДЕНИЯ КОРНЕЙ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ. УСКОРЕНИЕ СХОДИМОСТИ ПО ЗЙТКЕНУ

Наиболее употребительны для уточнения корней систем нелинейных уравнений – методы итерации (метод простой итерации и метод Зейделя) и метод Ньютона. Как и в случае уточнения корней одного нелинейного уравнения, для систем нелинейных уравнений требуется определение хорошего начального приближения (отделение корня), гарантирующего сходимость метода и высокую скорость сходимости. Для системы двух уравнений это может быть сделано графически, но для систем высоких порядков удовлетворительных методов отделения корней не существует.

Метод простой итерации. Систему нелинейных уравнений запишем в векторной форме

$$f(x) = 0, \quad (1.25)$$

где $x = (x_1, x_2, \dots, x_n)^T$ – вектор-столбец неизвестных (символ «Т» означает операцию транспонирования), $f(x) = [f_1(x), f_2(x), \dots, f_n(x)]^T$ – вектор-столбец функций. В методе простой итерации система (1.25) приводится к эквивалентной системе вида $x = \varphi(x)$, где $\varphi(x) = [\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)]^T$, или

$$\begin{cases} x_1 = \varphi_1(x_1, x_2, \dots, x_n), \\ x_2 = \varphi_2(x_1, x_2, \dots, x_n), \\ \dots \\ x_n = \varphi_n(x_1, x_2, \dots, x_n). \end{cases} \quad (1.26)$$

Полагая известным начальное приближение для корня $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$, построим итерационный процесс $x^{(k+1)} = \varphi(x^{(k)})$, $k = 0, 1, 2, \dots$, или

$$\begin{cases} x_1^{(k+1)} = \varphi_1[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}], \\ x_2^{(k+1)} = \varphi_2[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}], \\ \dots \\ x_n^{(k+1)} = \varphi_n[x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}]. \end{cases} \quad (1.27)$$

Рассмотрим поведение вектора погрешности

$$\varepsilon^{(k)} = x^{(k)} - x^* = [x_1^{(k)} - x_1^*, \dots, x_n^{(k)} - x_n^*],$$

полагая, что погрешность – малая величина. Для компонент вектора x можем записать

$$x_i^{(k+1)} = x_i^* - \varepsilon_i^{(k+1)} = \varphi_i[x_1^* - \varepsilon_1^{(k)}, \dots, x_n^* - \varepsilon_n^{(k)}], \quad i = 1, \dots, n.$$

Полагая наличие у функций $\varphi_i(x)$ непрерывных частных производных и используя $x_i^* = \varphi_i(x^*)$, можем (см. (1.19)), используя разложение в ряд, получить

$$\varepsilon_i^{(k+1)} \approx \sum_{j=1}^n \varepsilon_j^{(k)} \frac{\partial \varphi_i(x)}{\partial x_j} \Big|_{x=x^*}, \quad i=1, \dots, n. \quad (1.28)$$

Из (1.28) следует, что в методе простой итерации вектор погрешности испытывает линейное преобразование, или, иначе, метод имеет первый порядок сходимости.

Систему (1.28) можно переписать в виде

$$\varepsilon^{(k+1)} = A_\varphi \varepsilon^{(k)}, \quad (1.29)$$

где $A_\varphi = \begin{bmatrix} \frac{\partial \varphi_1(x)}{\partial x_1} & \frac{\partial \varphi_1(x)}{\partial x_2} & \frac{\partial \varphi_1(x)}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial \varphi_n(x)}{\partial x_1} & \frac{\partial \varphi_n(x)}{\partial x_2} & \frac{\partial \varphi_n(x)}{\partial x_n} \end{bmatrix}_{x=x^*}$ – матрица производных системы

функций $\varphi_i(x)$, $i=1, \dots, n$ (матрица Якоби). Достаточное условие сходимости итерационного процесса (1.27) формулируется следующим образом: если какая-либо норма матрицы A_φ , согласованная с рассматриваемой нормой вектора x , меньше единицы, то метод итераций сходится. Условие сходимости $\|A_\varphi\| < 1$ есть обобщение на случай нелинейной системы условия (1.20) для одного уравнения.

Метод итераций Зейделя. Нередко сходимость метода простой итерации можно улучшить, если вновь вычисленные значения компонент вектора неизвестных немедленно включить в расчет. В этом случае итерационный процесс имеет вид

$$x_i^{(k+1)} = \varphi_i[x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}, x_i^{(k)}, \dots, x_n^{(k)}], \quad i=1, \dots, n.$$

Сходимость этого процесса также линейная. Так же, как и при решении систем линейных уравнений, может быть поставлена задача об отыскании оптимальной на каждой итерации последовательности уточнения компонент вектора решения. Удовлетворительных методов построения оптимальной последовательности нет. На практике иногда используется упорядочение неизвестных по убыванию разности их значений на двух последовательных итерациях.

Метод Ньютона. Основная идея метода Ньютона – решение системы нелинейных уравнений $f(x)=0$ – сводится к решению последовательности линейных задач, дающих в пределе решение исходной задачи. Линейная задача получается путем выделения из нелинейных уравнений главной линейной части.

Рассмотрим погрешность вычисления корня на k -й итерации $\varepsilon^{(k)} = x^* - x^{(k)}$, где $\varepsilon^{(k)} = [\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \dots, \varepsilon_n^{(k)}]^T$. Полагая, что функции f_i непрерывны и дифференцируемы в окрестности корня и $\varepsilon_i^{(k)}$ ($i=1, \dots, n$) – малые величины, разложим $f(x^*) = f(x^{(k)} + \varepsilon^{(k)}) = 0$ в ряд Тейлора, сохранив лишь линейную часть разложения. Получим систему уравнений

$$\sum_{j=1}^n \varepsilon_j^{(k)} \frac{\partial f_i(x)}{\partial x_j} \Big|_{x=x^*} = -f_i(x^{(k)}), \quad i=1, \dots, n, \quad (1.30)$$

линейную относительно компонент вектора погрешностей. Если использовать эту систему для отыскания компонент вектора погрешностей, то в силу приближенности системы (1.30) (оставлена лишь линейная часть) найденное значение вектора погрешности будет лишь приближенным. Тогда при подстановке полученного решения в соотношение $x^* = \varepsilon^{(k)} - x^{(k)}$ будем иметь вместо x^* приближенное уточненное значение корня, которое обозначим $x^{(k+1)}$. Используя запись системы (1.30) в виде

$$A_f \varepsilon^{(k)} = -f^{(k)},$$

где A_f – матрица производных системы функций f_i (матрица Якоби), можем записать итерационный процесс для нахождения вектора x :

$$x^{(k+1)} = x^{(k)} - [A_f^{(k)}]^{-1} f^{(k)}, \quad k=0, 1, 2, \dots,$$

где $[A_f^{(k)}]^{-1}$ – матрица, обратная матрице Якоби. Представленная выше формула есть обобщение формулы (1.22) на случай систем нелинейных уравнений. Уточненное значение вектора может быть вновь использовано для получения следующего приближения к корню, что приводит к итерационному процессу. Заметим, что в большинстве случаев предпочтительным является не вычисление обратной матрицы $[A_f^{(k)}]^{-1}$, а получение каким-либо методом решения $\varepsilon^{(k)}$ линейной системы (1.30) и вычисление нового приближенного значения по соотношению $x^{(k+1)} = x^{(k)} + \varepsilon^{(k)}$.

Итерационный процесс (1.30) сходится, если определитель матрицы $A_f^{(k)}$ отличен от нуля, т.е. $\det(A_f^{(k)}) \neq 0$. Требуется, однако, хорошее отделение корня, но достаточное условие сходимости метода слишком громоздко, чтобы им можно было воспользоваться на практике.

На каждой итерации метода Ньютона требуется вычислять матрицу производных $A_f^{(k)}$ и решать систему линейных уравнений (1.30). Можно попытаться уменьшить объем вычислений за счет отказа от вычисления матрицы $A_f^{(k)}$ на

каждой итерации и использования на всех итерациях постоянного значения $A_f^{(0)}$, вычисленного по начальному приближению. Напомним, что при этом может быть дополнительно существенно уменьшен объем вычислений, если для решения последовательности линейных систем использовать алгоритм, позволяющий выполнить преобразование матрицы $A_f^{(0)}$ к верхней треугольной только один раз. Следует иметь в виду, однако, что, во-первых, указанная модификация метода Ньютона гарантирует лишь линейную сходимость итераций (против квадратичной в окрестности корня в методе Ньютона) и, во-вторых, константа в линейной зависимости погрешности при неудачном выборе начального приближения может оказаться весьма большой и сходимость будет медленной. Таким образом увеличивается число итераций, необходимое для достижения заданной точности, и уменьшение общего объема вычислений не гарантировано.

Ускорение сходимости по Эйткуэну. Предположим, что отношение $\frac{x^{(k-1)} - x^*}{x^{(k)} - x^*} = q$ есть величина постоянная и неизменна в процессе итераций. Тогда $\frac{x^{(k-1)} - x^*}{x^{(k)} - x^*} = q = \frac{x^{(k-2)} - x^*}{x^{(k-1)} - x^*}$. Отсюда имеем $x^* = \frac{x^{(k)}x^{(k-2)} - [x^{(k-1)}]}{x^{(k)} - 2x^{(k-1)} + x^{(k-2)}}$. Полученное таким образом x^* можно принять за следующее приближенное значение $x^{(k+1)}$. Предложенный способ пригоден как для одного нелинейного уравнения, так и для систем нелинейных уравнений. Это предположение означает, что метод применим к процессам с линейной сходимостью (простые итерации), но неприменим к методам Ньютона, секущих, парабол и т.п.

1.15. ВВЕДЕНИЕ В ПРОБЛЕМУ СОБСТВЕННЫХ ЗНАЧЕНИЙ

Во многих задачах механики, физики, химии и связанных с ними задачах алгебры и вычислительной математики вместе с квадратной матрицей A порядка n приходится рассматривать уравнение

$$\det(A - \lambda E) = 0, \quad (1.31)$$

называемое характеристическим или вековым уравнением матрицы A , и однородную систему

$$(A - \lambda E)x = 0, \quad (1.32)$$

где E – единичная матрица порядка n , x – вектор-столбец.

Уравнение (1.31) есть алгебраическое уравнение (многочлен) степени n вида

$$(-1)^n (\lambda^n - p_1 \lambda^{n-1} - p_2 \lambda^{n-2} - \dots - p_n) = (-1)^n P_n(\lambda). \quad (1.33)$$

Корни этого многочлена $\lambda_i, i=1, \dots, n$ называются собственными (характеристическими) значениями (числами) матрицы A . Однородная система (1.32)

имеет ненулевое решение лишь в том случае, если λ_i есть собственные значения матрицы A . Соответствующие каждому λ_i ненулевые решения x системы (1.32) называются собственными векторами матрицы A .

С точки зрения линейной алгебры задача о собственных значениях и векторах включает: составление характеристического многочлена (1.33), отыскание корней этого многочлена (собственных значений), решение для каждого λ_i однородной системы (1.32) – определение собственных векторов матрицы A .

Говорят о полной проблеме собственных значений, если требуется определить все собственные значения (и собственные векторы) матрицы A , и о частных проблемах собственных значений, если требуется найти лишь отдельные собственные значения, например, максимальное или минимальное. Обобщенная проблема собственных значений возникает при решении системы $Ax = \lambda Bx$, а общая – тогда, когда все коэффициенты матрицы A зависят от λ .

Существуют три группы методов решения второй проблемы линейной алгебры. В первой группе методов (методы отражения, вращения (прямой и итерационный) и QL-алгоритм) матрица приводится к какому-либо специальному виду, например диагональному или трехдиагональному, когда вычисление собственных чисел – простая задача. Во второй группе (методы Данилевского, Лерверье, Крылова, Ланцоша) матрицы приводятся к такой форме, чтобы вычисление коэффициентов характеристического многочлена было тривиальным, т.е. производится восстановление характеристического многочлена. И, наконец, в третьей группе – методов интерполяции – собственные числа определяются с использованием интерполяционных многочленов.

1.16. ИТЕРАЦИОННЫЙ МЕТОД ВРАЩЕНИЯ ДЛЯ НАХОЖДЕНИЯ СОБСТВЕННЫХ ЗНАЧЕНИЙ

Метод предложен в 1846 году Якоби и позволяет решить полную проблему собственных значений для эрмитовых матриц. Для простоты рассмотрим метод вращений для частного случая эрмитовых матриц – действительных симметрических матриц. В основе метода лежит тот факт, что действительная симметрическая матрица может быть приведена подобным преобразованием к диагональному виду. Метод Якоби основан на использовании нескольких оригинальных идей. Во-первых, используется преобразование подобия $S^{-1}AS$, сохраняющее собственные числа. Во-вторых, при этом преобразовании используются ортогональные матрицы, приводящие матрицы A к диагональной матрице Λ . Отметим, что априори соответствующая ортогональная матрица неизвестна и цель метода – в конечном итоге ее получить. В-третьих, согласно соотношению $\Lambda = U^{-1}AU$, в качестве ортогональной матрицы в каждой итерации используется матрица вращения. В-четвертых, итерационный процесс строится таким образом, чтобы после каждой итерации происходило уменьшение суммы квадратов внедиагональных элементов. В-пятых, реализация такого итерационного

процесса на каждом этапе приводит к обнулению максимального элемента матрицы и определению угла φ в матрице вращения.

Матрицей элементарного вращения называется матрица вида (она ортогональна)

$$U_{ij} = \begin{bmatrix} 1 & 0 & \dots & & & & & & \\ 0 & 1 & \dots & & & & & & \\ \dots & & \dots & & & & & & \\ 0 & 0 & \dots & \cos \varphi & \dots & -\sin \varphi & \dots & 0 & 0 \leftarrow i \\ \dots & & & & & & & & \\ 0 & 0 & \dots & & 1 & \dots & & 0 & 0 \\ \dots & & & & & & & & \\ 0 & 0 & \dots & \sin \varphi & \dots & \cos \varphi & \dots & 0 & 0 \leftarrow j \\ \dots & & & & & & & & \\ 0 & 0 & \dots & & & & & 0 & 1 \end{bmatrix}$$

$\begin{matrix} & & & \uparrow & & \uparrow & & \\ & & & i & & j & & \end{matrix}$

Название «матрица вращения» связано с тем, что в пространстве R^2 (на плоскости) матрица является матрицей преобразования декартовых координат при повороте осей на угол φ . Непосредственным умножением нетрудно убедиться, что $U_{ij}^{-1} = U_{ij}^T$, ибо, если матрица $U_{ij}(\varphi)$ является ортогональной, обратная матрица равна транспонированной. Столь простая возможность получения обратной матрицы, необходимой для выполнения подобного преобразования, весьма ценна с точки зрения эффективности метода. В итерационном методе вращения строится бесконечная последовательность подобных преобразований с помощью матриц вращения – такая, что в пределе исходная матрица преобразуется к диагональному виду. Рассмотрим алгоритм метода и покажем, что он действительно обеспечивает требуемое преобразование.

Предположим, что на некоторой k -й итерации построена матрица $A^{(k)}$. Выбираем в ней какой-либо внедиагональный элемент $a_{ij}^{(k)}$, максимальный по модулю. Так как матрица симметрична, то без ущерба для общности можно считать, что $i < j$. По индексам i и j строим матрицу вращения $U_{ij}^{(k)}$ и выполняем подобное преобразование над матрицей $A^{(k)}$. В результате получаем матрицу $A^{(k+1)} = [U_{ij}^{(k)} A^{(k)}]$, отличающуюся от матрицы $A^{(k)}$ только элементами i и j строк и столбцов.

Соотношения для вычисления элементов i и j строк и столбцов запишутся в следующем виде (для упрощения записи введем промежуточную матрицу $(B^{(k)} = [AU_{ij}^{(k)}])$). Учтывая, что вид матрицы $U_{ij}(\varphi)$, отметим, что все столбцы

матрицы $B^{(k)}$, кроме i и j , будут такими же, как и у матрицы $A^{(k)}$. Для элементов i и j столбцов этой матрицы

$$\begin{aligned} b_{mi}^{(k)} &= [a_{mi} \cos \varphi + a_{mj} \sin \varphi]^{(k)} \\ b_{mj}^{(k)} &= [-a_{mi} \sin \varphi + a_{mj} \cos \varphi]^{(k)}, \quad m = 1, 2, \dots, n. \end{aligned} \quad (1.34)$$

Заметим, что для всех $m \neq i$ и $m \neq j$, т.е. для внедиагональных элементов, можем записать $b_{mi} = a_{mi}^{(k+1)}$, так как последующее умножение матрицы B слева на U_{ij}^T не изменяет внедиагональных элементов i и j столбцов.

Для элементов i и j строк матрицы $A^{(k+1)}$

$$\begin{aligned} a_{im}^{(k+1)} &= [b_{im} \cos \varphi + b_{jm} \sin \varphi]^{(k)} \\ a_{jm}^{(k+1)} &= [-b_{im} \sin \varphi + b_{jm} \cos \varphi]^{(k)}, \quad m = 1, 2, \dots, n. \end{aligned} \quad (1.35)$$

Матрицу элементарного вращения $U_{ij}^{(k)}$ выбираем таким образом, чтобы обратить максимальный по модулю внедиагональный элемент в нуль. Для этого, используя соотношения (1.34) и (1.35), можем записать

$$a_{ij}^{(k+1)} = \left[\frac{1}{2} (a_{jj} - a_{ii}) \sin 2\varphi + \cos 2\varphi \right]^{(k)}.$$

Приравняв $a_{ij}^{(k+1)}$ нулю, получаем выражение для определения угла $\varphi^{(k)}$, обеспечивающего при подобном преобразовании с помощью матрицы вращения зануление элемента матрицы $A^{(k+1)}$ с индексами i и j и симметричного ему элемента. Имеем

$$\operatorname{tg} 2\varphi^{(k)} = \left[\frac{2a_{ij}^{(k)}}{(a_{ii}^{(k)} - a_{jj}^{(k)})} \right]^{(k)}. \quad (1.36)$$

Таким образом, подобное преобразование при выборе матрицы вращения с соотношением (1.36) зануляет элемент с индексами i и j и симметричный ему элемент матрицы $A^{(k+1)}$. При этом, однако, в соответствии с соотношением (1.34) и (1.35) происходит изменение всех остальных элементов i и j строк и столбцов (и, в частности, тех, которые, может быть, на предыдущих итерациях были обращены в нуль). Показывается, что в построенном процессе все внедиагональные элементы матрицы при достаточном числе подобных преобразований (итераций) становятся сколь угодно малыми.

Можно показать, что сумма квадратов внедиагональных элементов $t(A)$ на каждой итерации уменьшается, так что имеет место соотношение

$$t[A^{(k+1)}] = t[A^{(k)}] - 2[a_{ij}^{(k)}]^2 \quad \text{или} \quad t[A^{(k+1)}] < t[A^{(k)}].$$

Можно также показать, что $t[A^{(k)}] = q^k t(A)$, где $q < 1 - 2/(n(n+1))$, т.е. сходимость итерационного метода вращения ухудшается с увеличением n .

Собственные векторы являются столбцами матрицы $U = \prod_{k=0}^K U_y^{(k)}$. В методе вращения для вычисления собственных значений требуется $50n^3$ операций. В качестве хорошего нулевого приближения к методу Гаусса служит метод вращения Гивенса, применимый к неэрмитовым матрицам.

1.17. МЕТОД ДАНИЛЕВСКОГО ДЛЯ ПОСТРОЕНИЯ ХАРАКТЕРИСТИЧЕСКОГО МНОГОЧЛЕНА МАТРИЦЫ

В методе Данилевского исходная матрица A с помощью подобных преобразований $\Phi = S^{-1}AS$ приводится к так называемой канонической форме Фробениуса:

$$\Phi = \begin{bmatrix} p_1 & p_2 & p_3 & \dots & p_{n-1} & p_n \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}. \quad (1.37)$$

Разложив определитель матрицы $(\Phi - \lambda E)$ по элементам первой строки, получим

$$|\Phi - \lambda E| = (-1)^n (\lambda^n - p_1 \lambda^{n-1} - p_2 \lambda^{n-2} - \dots - p_n) = (-1)^n P_n(\lambda),$$

т.е. элементы первой строки матрицы (1.37) являются коэффициентами характеристического многочлена матрицы Φ , и так как матрица Φ получена подобными преобразованиями матрицы A , то они являются также и коэффициентами характеристического многочлена исходной матрицы A .

Переход от матрицы A к матрице Φ осуществляется за $(n-1)$ преобразование подобия, последовательно преобразующее строки матрицы A , начиная с последней, в строки матрицы Φ . Рассмотрим схему преобразований. Пусть после некоторого шага матрица A приведена к виду

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1(k-1)} & b_{1k} & \dots & b_{1(n-1)} & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2(k-1)} & b_{2k} & \dots & b_{2(n-1)} & b_{2n} \\ & & \dots & & & & & \\ b_{k1} & b_{k2} & \dots & b_{k(k-1)} & b_{kk} & \dots & b_{k(n-1)} & b_{kn} \\ 0 & 0 & & 0 & 1 & \dots & 0 & 0 \\ & & & & & \dots & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(здесь и далее индекс шага для сокращения опущен).

Приведение k -й строки к строке матрицы Фробениуса осуществляется следующим образом. Весь $(k-1)$ -й столбец делится на $b_{k(k-1)}$, так что вместо элемента $b_{k(k-1)}$ имеем 1. Далее этот столбец умножается на элемент $-b_{1k}$ и складывается с i -м столбцом, при этом i меняется от 1 до n (кроме, очевидно, $(k-1)$). В результате k -я строка принимает форму Фробениуса. Нетрудно показать, что описанная процедура эквивалентна умножению матрицы B справа на матрицу S , которая имеет вид

$$S = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ & & \dots & & & \\ -\frac{b_{k1}}{b_{k(k-1)}} & -\frac{b_{k2}}{b_{k(k-1)}} & \dots & \frac{1}{b_{k(k-1)}} & \dots & -\frac{b_{kn}}{b_{k(k-1)}} \\ & & & \dots & & \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}.$$

Непосредственным умножением легко убедиться, что обратная S матрица S^{-1} имеет вид

$$S^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ & & \dots & & & \\ b_{k1} & b_{k2} & \dots & b_{kk} & \dots & b_{kn} \\ & & & \dots & & \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}.$$

При умножении матрицы B справа на матрицу S k -я строка матрицы B принимает требуемый вид, а строки матрицы Фробениуса с $(k-1)$ по n сохраняют свой вид.

Однако матрица BS не подобна матрице A , и чтобы сделать её подобной, нужно BS умножить слева на S^{-1} .

Умножение матрицы BS слева на S^{-1} не меняет строк с k -й по n -ю этой матрицы. Применяя последовательно описанный процесс к строкам матрицы A , начиная с последней, получим матрицу Фробениуса (1.37).

1.18. МЕТОД ИНТЕРПОЛЯЦИИ ДЛЯ ПОСТРОЕНИЯ ХАРАКТЕРИСТИЧЕСКОГО МНОГОЧЛЕНА

В основе метода интерполяции для построения характеристического многочлена лежит известная теорема алгебры о единственности интерполяцион-

ного многочлена. Задача об определении коэффициентов $p_i, i = 1, \dots, n$, характеристического многочлена степени n сводится к задаче построения интерполяционного многочлена степени n по известным значениям его в узловых точках (коэффициент при неизвестном в степени n в (1.33) равен $(-1)^n$). Для этого задается n различных значений $\lambda = \beta_j, j = 1, 2, \dots, n$, и для каждого из этих значений вычисляется значение определителя $\det(A - \beta_j E) = \alpha_j$. Значения β_j могут выбираться произвольно, но для уменьшения влияния ошибок округления желательно узлы интерполяции располагать равномерно и не слишком выходя за границы интервала $(\lambda_{\min}, \lambda_{\max})$. Для оценки границ интервала можно использовать соотношение $\max |\lambda| \leq \|A\|$, где $\|A\|$ – любая норма матрицы A .

Для определения $p_i (i = 1, \dots, n)$ записывается система n линейных уравнений

$$\beta_j^n - \sum_{i=1}^n p_i \beta_j^{n-i} = \alpha_j, j = 1, \dots, n.$$

Если выбранные значения β_j различны, то определитель этой системы отличен от нуля и система уравнений имеет единственное решение. В силу теоремы о единственности интерполяционного многочлена построенный таким образом многочлен является характеристическим многочленом матрицы A . Отметим, что обычно для построения многочлена используются интерполяционные формулы, которые будут рассмотрены далее.

Одним из существенных достоинств метода интерполяции является его общность: метод может быть использован для построения характеристического многочлена для матриц произвольной структуры. Вместе с тем общность метода является и его недостатком, так как не позволяет учесть некоторые особенности матриц, возникающих в практических задачах.

Для применения метода интерполяции требуется вычислить n определителей порядка n . Вместе с построением интерполяционного многочлена по интерполяционной формуле метод требует примерно n^4 операций, что приводит для больших n к довольно большому объему вычислений. Однако более серьезным обстоятельством, ограничивающим применение метода интерполяции, является то, что нули многочлена (т.е. искомые собственные значения) могут быть весьма чувствительны к погрешностям коэффициентов многочлена. Незбежные и, возможно, достаточно большие погрешности вычисления коэффициентов P_n в методе интерполяции приводят к существенному снижению точности определения собственных значений при больших n . По этой причине метод интерполяции редко используется при $n > 10$.

Последний недостаток метода интерполяции можно устранить, если использовать метод локальной интерполяции. В этом методе восстанавливается не весь характеристический многочлен, а некоторая его часть вблизи корня. Вычислив, например, $\det(A - \lambda E)$ в трех точках, можно построить в окрестности

корня многочлен второй степени и далее, решая квадратное уравнение, определить корни. Для нахождения корня используется метод парабол. Метод локальной интерполяции обладает тем преимуществом, что избавляет от необходимости строить многочлен высших степеней. Кроме того, он пригоден для матриц любых видов, а также может быть использован для решения частичной, полной, обобщенной и общей проблемы.

В методах интерполяции, таким образом, решение второй проблемы достигается путем многократного решения первой проблемы линейной алгебры, имея в виду вычисление определителей.

Глава 2. ПРИБЛИЖЕНИЕ ФУНКЦИЙ. ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ И ДИФФЕРЕНЦИРОВАНИЕ

2.1. КРАТКОЕ СОДЕРЖАНИЕ ГЛАВЫ

Материал этой главы объединяет общий подход к решению задач. Сначала излагается интерполяция и строятся интерполяционные многочлены в форме Лагранжа и Ньютона; из методов среднеквадратичной аппроксимации рассматривается метод наименьших квадратов. Затем рассмотрена сплайн-интерполяция. Формулы численного дифференцирования получены дифференцированием интерполяционного многочлена и разложением в ряд Тейлора. Обсуждается точность формул и пути повышения точности численного дифференцирования. Излагаемые затем формулы численного дифференцирования получены на основе многочленной интерполяции. Исследованы точность и характер сходимости этих формул, рассмотрена возможность уточнения результата численного интегрирования (метод Рунге-Ромберга).

2.2. ПОСТАНОВКА ЗАДАЧИ О ПРИБЛИЖЕНИИ (АППРОКСИМАЦИИ) ФУНКЦИЙ

Во многих практически важных задачах возникает проблема замены некоторой функции, заданной аналитически или таблично, другой, близкой к первой, в некотором смысле, но более простой и удобной при вычислениях. Например, замена функции многочленом позволяет получить простые формулы численного интегрирования и дифференцирования; замена таблицы приближающей функцией позволяет получить значения в её промежуточных точках. Эти примеры можно продолжить и дальше.

Возникает также и вторая задача – восстановление функции на некотором отрезке по заданным на этом отрезке значениям функции в дискретном множестве точек. Отметим также, что теория приближения функций является важным вспомогательным аппаратом при численном решении дифференциальных уравнений.

В общем случае при постановке задачи приближения необходимо ответить на следующие вопросы:

во-первых, какой класс приближенных функций необходимо выбрать. Ответ на такой вопрос зависит от вида приближаемой функции и её целей, для которых в дальнейшем будет использоваться приближающая функция. Широко используются следующие классы функций: многочлены, тригонометрические функции, показательные функции и др.;

во-вторых, надо выбрать критерий близости исходной и приближающей функции. В качестве критерия можно выбрать, например, точное совпадение приближаемой и приближающей функций в узловых точках (лагранжева интерполяция); минимум суммы квадратов отклонения в узловых точках (метод наименьших квадратов) и другие. Как и при выборе класса приближающих функций, выбор критерия близости исходной и приближающей функций определяется целью построения приближающей функции и может сильно повлиять на результаты. При аппроксимации экспериментальных результатов целесообразно использовать среднеквадратичное приближение. На рис. 2.1 показаны эти два случая приближения функций. В первом случае целесообразно использовать интерполяцию (сплошная линия), во втором – среднеквадратичное приближение (пунктир), так как интерполяция (сплошная линия) может неверно описывать свойства функций и лишь усугублять ошибки эксперимента;

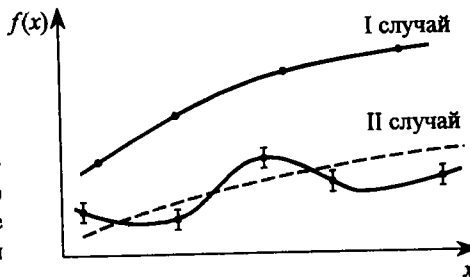


Рис. 2.1

в-третьих, необходимо указать правило, позволяющее с заданной степенью точности получить значение функции в промежутках между узлами, в частности, ответить на вопросы, какие узлы использовать для построения приближающей функции и как их расположить

Таким образом, видим, что построение приближающей функции существенно зависит от ответа на перечисленные выше вопросы.

2.3. ИНТЕРПОЛЯЦИЯ

Пусть на отрезке $[a, b]$ задано дискретное множество несовпадающих точек x_i , которые будем называть узлами и в которых известны значения функции $f_i = f(x_i)$, $i = 0, \dots, n$. Потребуем, чтобы приближающая функция $\varphi(x, a)$ совпадала с приближаемой $f(x)$ в $(n+1)$ узле таблицы, т.е. потребуем выполнения равенства

$$\varphi(x_i, a_0, \dots, a_n) = f(x_i) = f_i, \quad i = 0, \dots, n. \quad (2.1)$$

Этот способ построения приближающей функции, при котором в узлах значения приближаемой и приближающей функций совпадают, называется интерполяцией, или лагранжевой интерполяцией. Наиболее распространен способ линейной интерполяции, в случае которой приближающая функция ищется в виде линейной комбинации некоторых базисных функций $\varphi_i(x)$:

$$\varphi(x_i, a_0, \dots, a_n) = \sum_{i=0}^n a_i \varphi_i(x). \quad (2.2)$$

Система функций $\varphi_i(x)$ должна быть линейно независимой, так как в противном случае число параметров и членов в сумме можно было бы уменьшить и, кроме того,

$$\det \begin{vmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \dots & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \dots & \varphi_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_0(x_n) & \varphi_1(x_n) & \dots & \varphi_n(x_n) \end{vmatrix} \neq 0.$$

Подставляя (2.2) в (2.1), получаем систему линейных уравнений для определения коэффициентов a_i :

$$f_k = \sum_{i=0}^n a_i \varphi_i(x_k), \quad k = 0, \dots, n.$$

В качестве базисных функций можно выбрать любую линейно независимую систему функций, но чаще всего выбираются степенные функции $1, x, x^2, \dots, x^n$. Это объясняется тем, что многочлены легко вычисляются и теория интерполяции многочленами хорошо разработана. В случае приближения многочленами приближающую функцию ищем в виде многочлена степени n :

$$P_n(x) = \sum_{i=0}^n a_i x^i, \quad (2.3)$$

где нижний индекс n указывает на степень интерполяционного многочлена. Подставляя в (2.3) значения узлов и используя условие $P_n(x_i) = y_i$, получаем систему линейных алгебраических уравнений относительно коэффициентов a_i :

$$\sum_{i=0}^n a_i x^i = f_k, \quad k = 0, \dots, n. \quad (2.4)$$

Определитель системы (2.4) является в случае несовпадения узлов интерполяции отличным от нуля определителем Вандермонда:

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \prod_{j>i \geq 0} (x_j - x_i) \neq 0.$$

Таким образом, решение системы (2.4) существует и единственно. А это значит, что интерполяционный многочлен $P_n(x)$ существует и единствен с точностью до формы записи.

2.4. ИНТЕРПОЛЯЦИОННЫЙ МНОГОЧЛЕН В ФОРМЕ ЛАГРАНЖА

Будем искать интерполяционный многочлен в виде

$$L_n(x) = \sum_{i=0}^n f_i l_i(x). \quad (2.5)$$

Здесь $l_i(x)$ – многочлены степени n , так называемые лагранжевы многочлены влияния, удовлетворяющие условию

$$l_i(x_j) = \begin{cases} 1, & \text{при } i = j, \\ 0, & \text{при } i \neq j. \end{cases}$$

Последнее условие означает, что любой многочлен $l_i(x)$ равен нулю при каждом x_j кроме x_i , т.е. $x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ – корни этого многочлена. Следовательно, $l_i(x)$ имеют вид

$$l_i(x) = C_i (x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n).$$

Так как по условию $l_i(x_i) = 1$, то

$$C_i = 1 / \{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)\}.$$

Таким образом, лагранжевы многочлены влияния

$$l_i(x) = \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)},$$

а интерполяционный многочлен (2.5) запишется в виде

$$L_n(x) = \sum_{i=0}^n f_i \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)}. \quad (2.6)$$

Интерполяционный многочлен, записанный в форме (2.6), называется интерполяционным многочленом Лагранжа. Важным достоинством этой формы записи интерполяционного многочлена является то, что число арифметических операций, необходимых для построения многочлена Лагранжа, пропорционально n^2 и является наименьшим для всех форм записи. К достоинствам можно отнести также и то, что формула (2.6) в явном виде содержит значения функций в узлах интерполяции, что бывает удобно при некоторых вычислениях, в частности при построении формул численного интегрирования. Формула (2.6) применима как для равноотстоящих, так и для неравноотстоящих узлов. Достоинством интерполяционного многочлена Лагранжа является и то, что он особенно удобен, когда значения функций меняются, а узлы интерполяции неизменны,

что имеет место во многих экспериментальных исследованиях. К недостаткам этой формы записи можно отнести то, что с изменением числа узлов приходится все вычисление проводить заново. Это затрудняет проведение апостериорных оценок точности (оценок, получающихся в процессе расчета).

Введем функцию $\omega_{n+1} = (x - x_0)(x - x_1) \dots (x - x_n) = \prod_{i=0}^n (x - x_i)$.

Отметим, что $\omega_{n+1}(x)$ есть многочлен степени $n + 1$. Тогда формулу (2.6) можно записать в виде

$$L_n(x) = \sum_{i=0}^n f_i \frac{\omega_{n+1}(x)}{(x - x_i) \omega'_{n+1}(x_i)}. \quad (2.7)$$

Приведем формулы линейной и квадратичной интерполяции по Лагранжу

$$L_1(x) = f_0 \frac{(x - x_1)}{(x_0 - x_1)} + f_1 \frac{(x - x_0)}{(x_1 - x_0)}, \quad (2.8)$$

$$L_2(x) = f_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}. \quad (2.9)$$

В (2.8) и в (2.9) многочлен Лагранжа имеет первую и вторую степень соответственно.

Эти формулы наиболее часто используются на практике. Пусть задан $(n + 1)$ узел интерполяции. На этих узлах можно построить один интерполяционный многочлен n -й степени, $(n - 1)$ многочлен первой степени и большой набор многочленов степени меньше n , опирающиеся на некоторые из этих узлов. Теоретически максимальную точность обеспечивает многочлен более высокой степени. Однако на практике наиболее часто используют многочлены невысоких степеней, во избежание погрешностей при расчетах коэффициентов при больших степенях многочлена.

2.5. ИНТЕРПОЛЯЦИОННЫЙ МНОГОЧЛЕН В ФОРМЕ НЬЮТОНА

Интерполяционный многочлен в форме Ньютона представляет иную форму записи интерполяционного многочлена. Но так как эта форма записи имеет ряд достоинств по сравнению с многочленом Лагранжа, получим её.

Для вывода введем понятие разделенной разности. Разделенные разности нулевого порядка совпадают со значениями функции в узлах. Разделенные разности первого порядка обозначаются $f(x_i, x_j)$ и определяются через разделенные разности нулевого порядка:

$$f(x_i, x_j) = \frac{f_i - f_j}{x_i - x_j},$$

разделенные разности второго порядка определяются через разделенные разности первого порядка:

$$f(x_i, x_j, x_k) = \frac{f(x_i, x_j) - f(x_j, x_k)}{x_i - x_k}.$$

Разделенная разность порядка n определяется соотношениями

$$f(x_i, x_j, x_k, \dots, x_{n-1}, x_n) = \frac{f(x_i, x_j, x_k, \dots, x_{n-1}) - f(x_j, x_k, \dots, x_n)}{x_i - x_n}. \quad (2.10)$$

Таким образом, для $(n + 1)$ -й точки могут быть построены разделенные разности до n -го порядка; разделенные разности более высоких порядков равны нулю. Из вида разделенных разностей можно сделать заключение, что они являются аналогом производных соответствующих порядков.

Между разделенными разностями и производными соответствующих порядков существует соотношение $f^{(n)}(x) = n! f(x_0, x_1, \dots, x_n)$.

Очевидно, что разделенная разность $(n + 1)$ -го порядка от многочлена n -й степени равна нулю.

Во многих учебниках для таблиц с постоянным шагом h вводят конечные разности $\Delta^n f$, связанные с разделенными соотношением

$$\Delta^n f = h^n n! f(x_0, x_1, \dots, x_n).$$

Приведем вывод формулы интерполяционного многочлена Ньютона. Пусть аппроксимируемая функция $f(x)$ – многочлен n -й степени. Выберем точки $x_0, x_1, \dots, x_n$ в качестве узлов и найдем такой интерполяционный многочлен, значения которого в узлах интерполяции совпадают со значениями функции $f(x)$.

Запишем, выбирая в качестве узла точку с координатой x , очевидные соотношения:

$$f(x) - f(x_0) = (x - x_0)f(x, x_0), \quad (2.11)$$

$$f(x) = f(x_0) + (x - x_0)f(x, x_0).$$

Имеем также:

$$f(x, x_0) - f(x_0, x_1) = (x - x_1)f(x, x_0, x_1) \quad \text{или} \quad (2.12)$$

$$f(x, x_0) = f(x_0, x_1) + (x - x_1)f(x, x_0, x_1).$$

Подставляя (2.12) в (2.11), получаем

$$f(x) = f(x_0) + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x, x_0, x_1).$$

Далее:

$$f(x, x_0, x_1) - f(x_0, x_1, x_2) = (x - x_2)f(x, x_0, x_1, x_2) \quad \text{или} \quad (2.13)$$

$$f(x, x_0, x_1) = f(x_0, x_1, x_2) + (x - x_2)f(x, x_0, x_1, x_2).$$

Подставляя (2.13) в предыдущее выражение, имеем

$$f(x) = f(x_0) + (x - x_0)f(x_1, x_0) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + (x - x_0)(x - x_1) \times (x - x_2)f(x, x_0, x_1, x_2).$$

Продолжая процесс, получаем в итоге:

$$P_n(x) = f(x_0) + (x - x_0)f(x_1, x_0) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \dots + (x - x_0)(x - x_1) \dots (x - x_n)f(x_0, x_1, \dots, x_n). \quad (2.14)$$

Следующий член в этом ряду будет равен нулю, поскольку он должен содержать произведение $\omega_{n+1}(x)$ и разделенной разности $(n+1)$ -го порядка.

Так как $\omega_{n+1}(x)$ – многочлен степени $n+1$, а исходный многочлен имеет степень n , то разделенная разность $(n+1)$ -го порядка должна быть равна нулю. В силу условия единственности интерполяционного многочлена $P_n(x)$ есть интерполяционный многочлен n -й степени с узлами $x_0, x_1, \dots, x_n$.

Запись многочлена в формуле (2.14) есть так называемый интерполяционный многочлен Ньютона. Если функция $f(x)$ не есть многочлен n -й степени, то формула (2.14) для $P_n(x)$ имеет место, однако она не точно приближает функцию $f(x)$, а с некоторой погрешностью. Отметим, что при добавлении новых узлов первые члены многочлена Ньютона остаются неизменными, что видно из сравнения формул (2.11) – (2.14).

Если функция задана в точках $x_0, x_1, \dots, x_n$, то при построении интерполяционного многочлена Ньютона удобно пользоваться таблицей, называемой таблицей разделенных разностей, пример которой для $n=3$ приведен в табл. 2.1.

Таблица 2.1

x_0	$f(x_0)$			
x_1	$f(x_1)$	$f(x_0, x_1)$		
x_2	$f(x_2)$	$f(x_1, x_2)$	$f(x_0, x_1, x_2)$	
x_3	$f(x_3)$	$f(x_2, x_3)$	$f(x_1, x_2, x_3)$	$f(x_0, x_1, x_2, x_3)$

За точностью расчета удобно следить по убыванию членов суммы (2.14). Если величина членов убывает быстро, то оставляют только те из них, которые больше допустимой погрешности. Для повышения точности интерполяции в сумму могут быть добавлены новые члены, что требует подключения дополнительных узлов. При этом безразлично, в каком порядке подключаются новые узлы. Этим формула Ньютона выгодно отличается от формулы Лагранжа, при добавлении в которую новых узлов все расчеты надо производить заново. Интерполяционный многочлен Ньютона удобен при построении формул численного дифференцирования, что очевидно из его вида. Кроме того, он удобен для проведения апостериорных оценок.

Приведем формулы линейной и квадратичной интерполяции в соответствии с многочленом Ньютона. Имеем:

$$P_1(x) = f_0 + (x - x_0) \frac{f_1 - f_0}{x - x_1}; \quad (2.15)$$

$$P_2(x) = f_0 + (x - x_0) \frac{f_1 - f_0}{x_0 - x_1} + (x - x_0)(x - x_1) \frac{\frac{f_1 - f_0}{x_0 - x_1} - \frac{f_2 - f_1}{x_1 - x_2}}{x_0 - x_2}. \quad (2.16)$$

Нетрудно убедиться, что эти формулы будут совпадать с соответствующими формулами для интерполяционного члена Лагранжа после простых преобразований. Очевидно, что это утверждение есть следствие условия единственности интерполяционного многочлена.

2.6. ПОГРЕШНОСТЬ И СХОДИМОСТЬ ИНТЕРПОЛЯЦИИ

Построив для функции $f(x)$ интерполяционный многочлен $P_n(x)$ или $L_n(x)$, необходимо выяснить, насколько близко интерполяционный многочлен приближается к функции в других точках отрезка. Ошибку интерполяции можно представить в следующем виде:

$$R_n(x) = \varepsilon_n(x) = f(x) - P_n(x) = \omega_{n+1}(x)r(x), \quad (2.17)$$

так как ошибка интерполяции обращается в нуль в узлах сетки. Выражение для $r(x)$ найдем из следующих соображений. Введем функцию $q(x) = f(x) - P_n(x) - \omega_{n+1}(x)r(x^*)$, где x^* – произвольная точка. Очевидно, что $q(x) = 0$ в $(n+2)$ точках, в которых $x = x_i, i = 0, \dots, n$, и при $x = x^*$. Если $f(x)$ имеет $(n+1)$ -ю непрерывную производную, то и $q(x)$ можно продифференцировать $(n+1)$ раз. А так как между двумя нулями гладкой функции – нуль её производной, то $q'(x)$ обращается в нуль на всем интервале по крайней мере $(n+1)$ раз. Последовательно применяя это правило дальше, получаем, что между крайними нулями функции $q^{(n)}(x)$ лежит, согласно теореме Ролля, по крайней мере один нуль функции $q^{(n+1)}(x)$, т.е. существует такая точка ξ , принадлежащая отрезку интерполирования, в которой выполняется следующее равенство:

$$q^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n+1)! r(x^*) = 0.$$

Следовательно, $r(x^*) = f^{(n+1)}(\xi)/(n+1)!$. Так как x^* – произвольная точка, то погрешность интерполяции запишется следующим образом:

$$R_n(x) = \varepsilon_n(x) = f(x) - P_n(x) = \omega_{n+1}(x) \frac{f^{(n+1)}(\xi(x))}{(n+1)!}, \quad (2.18)$$

т.е. можно сказать, что существует точка ξ , для которой выполнено равенство (2.18). Так как точка ξ неизвестна, то обычно используют мажорантную оценку. Заменяя величины, входящие в (2.18), их максимальными значениями, получаем величину максимальной погрешности

$$|\varepsilon_n(x)| = |f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\omega_{n+1}(x)|, \quad (2.19)$$

где $M_{n+1} = \max |f^{(n+1)}(\xi)|$, $\xi \in [x_0, x_n]$.

Формула (2.19) дает возможность провести априорную оценку точности, т.е. для случая аналитически заданной функции $f(x)$ провести оценку до начала вычислений. Однако, так как величину производных интерполированной функции бывает оценить достаточно сложно (например, для таблично заданной функции), то априорной оценкой пользуются редко. Удобнее применять апостериорную оценку по первому отброшенному члену интерполяционного многочлена Ньютона. Это удобство является следствием того факта, что в процессе построения многочлена Ньютона вычисляются разделенные разности, являющиеся, как отмечалось выше, аналогами производных соответствующих порядков.

На рис. 2.2 представлена функция $\omega_{(n+1)}(x)$ для равноотстоящих узлов. Из рис. 2.2 видно (на примере интерполяционного многочлена пятой степени), что

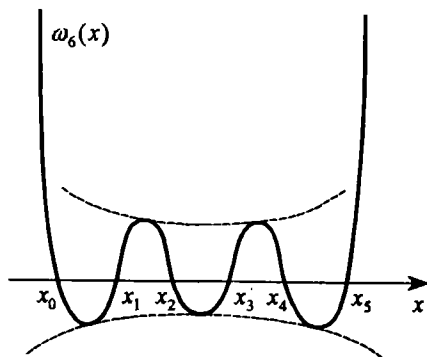


Рис. 2.2

максимумы $\omega_{(n+1)}(x)$ меньше вблизи центрального узла интерполяции, больше вблизи крайних узлов и быстро возрастают при выходе за центральные узлы интерполяции. Приближение, когда точка, в которой ищется значение приближающей функции, находится за пределами отрезка интерполяции, называется экстраполяцией. При экстраполяции в точках, далеко отстоящих от концов отрезка интерполяции, ошибка приближения может стать значительной. Ошибка интерполяции в центральном интервале

при постоянном шаге таблицы определяется следующим выражением:

$$|\varepsilon_n(x)| \leq \sqrt{\frac{2}{n}} M_{n+1} \left[\frac{h}{2} \right]^{n+1}. \quad (2.20)$$

Поэтому при постоянном шаге таблицы точку x выгодно выбирать вблизи середины отрезка интерполирования.

Из (2.20) видно, что если интерполяционный многочлен имеет порядок n , то ошибка интерполяции будет убывать как h^{n+1} при переходе от таблиц с более крупным шагом к таблицам с более мелким шагом. Следовательно, если производная, входящая в оценку погрешности, ограничена, то при фиксированных степени многочлена и изменении шага сетки погрешность $|f(x) - P_n(x)|$ неограниченно убывает и многочлен равномерно сходится к $f(x)$ на ограниченном отрезке. Если значение производной оценивается легко, то из (2.20) легко оценить число узлов, обеспечивающих заданную точность интерполяции.

Второй способ уменьшения ошибки интерполяции заключается в увеличении числа узлов, т.е. степени интерполяции, при постоянном шаге сетки. При этом знаменатель в (2.19) возрастает. Этот способ не всегда приводит к уменьшению ошибки за счет роста $(n+1)!$, так как функция f может не иметь $(n+1)$ -ю производную, кроме того, её максимум может возрастать очень быстро.

Трудности, возникающие при многочленной интерполяции, хорошо иллюстрируются примером Рунге: интерполяция на отрезке $[-1, 1]$ функции $f(x) = 1/(1+25x^2)$ при равномерном распределении узлов (рис. 2.3). С увеличением порядка n интерполяционного многочлена возрастают колебательные свойства многочлена, и это приводит к тому, что интерполяционный многочлен при $0.726 \leq x \leq 1$ расходится.

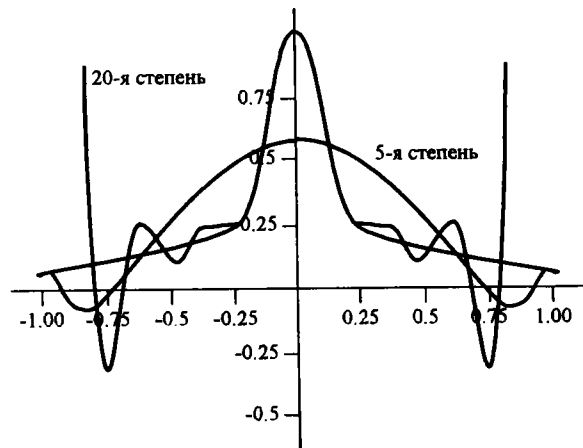


Рис. 2.3

Задача об оптимальном распределении узлов на отрезке интерполирования, приводящем к минимизации ошибки интерполирования, решена П.Л. Чебышевым. Им показано, что если узлы интерполяции не распределены равномерно, а совпадают с корнями так называемых чебышевских полиномов степе-

ни $n+1$, то ошибка многочленной интерполяции минимальна. При таком выборе узлов интерполяции проблема расходимости многочлена для процедуры Рунге исчезает. Если интерполяция производится на интервале (a,b) , то узлы интерполяционного многочлена следует расположить в нулях многочлена Чебышева, т.е. в точках x_i , где

$$x_i = \frac{b+a}{2} + \frac{b-a}{2} z_i,$$

$$z_i = \cos \left[\frac{\pi}{2(n+1)} + i \frac{\pi}{n+1} \right], \quad i = 0, \dots, n.$$

На практике для повышения точности интерполяционного многочлена, как правило, уменьшают шаг таблицы, оставляя неизменной степень интерполирующего многочлена. Для того чтобы уменьшить влияние колебательных свойств интерполяционного многочлена, пользуются линейной или квадратичной интерполяцией.

2.7. ИНТЕРПОЛЯЦИЯ СПЛАЙНАМИ

Использование одной интерполяционной формулы на большом числе узлов, как отмечалось в предыдущем разделе, нецелесообразно. Такой интерполяционный многочлен сильно проявляет свои колебательные свойства, и его значения между узлами могут сильно отличаться от значений интерполируемой функции. Одна из возможностей преодоления этого недостатка заключается в применении сплайн-интерполяции. Суть сплайн-интерполяции заключается в определении интерполирующей функции по формулам одного типа для различных подмножеств и в стыковке значений функции и её производных на границах подмножеств

Наиболее изученным и широко применяемым является случай, когда между любыми двумя точками строится многочлен n -й степени

$$S(x) = \sum_{k=0}^n a_{ik} x^k, \quad x_{i-1} \leq x \leq x_i,$$

который в узлах интерполяции принимает значения интерполируемой функции и непрерывен вместе со своими $(n-1)$ производными. Такой кусочно-непрерывный интерполяционный многочлен называется сплайном. Его коэффициенты находятся из условий в узлах сетки: равенства значений сплайна и приближаемой функции, а также равенства $n-1$ производной соответствующих многочленов.

Максимальная по всем частичным отрезкам степень многочленов называется степенью сплайна, а разница между степенью сплайна и порядком наимышей непрерывной производной – дефектом сплайна (def).

Одним из наиболее распространенных интерполяционных сплайнов является кубический интерполяционный сплайн (def = 1). Для вывода уравнения

кубического интерполяционного сплайна воспользуемся его интерпретацией как гибкой линейки, изогнутой таким образом, что она проходит через значения функции в узлах сетки, т.е. является упругим брусом в состоянии свободного равновесия. Так как это его состояние описывается уравнением $S^{IV}(x) = 0$, где $S^{IV}(x)$ – четвертая производная, то между каждой парой соседних узлов интерполяционная формула записывается в виде многочлена третьей степени. Этот многочлен удобно представить как

$$S(x) = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3,$$

$$x_{i-1} \leq x \leq x_i, \quad i = 1, 2, \dots, n.$$

Для построения кубического сплайна необходимо построить n многочленов третьей степени, т.е. определить $4n$ неизвестных a_i, b_i, c_i, d_i . Эти коэффициенты ищутся из условий в узлах сетки. По определению, в узлах сетки сплайн (2.21) должен принимать табличные значения функции:

$$S(x_{i-1}) = a_i = f_{i-1}, \quad (2.22)$$

$$S(x_i) = a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 = f_i,$$

$$h_i = x_i - x_{i-1}. \quad (2.23)$$

Система (2.22), (2.23) содержит $2n$ уравнений. Дополнительные уравнения можно получить, если потребовать непрерывности первой и второй производной функции $S(x)$ во внутренних узлах сетки:

$$S'(x) = b_i + 2c_i(x - x_{i-1}) + 3d_i(x - x_{i-1})^2,$$

$$S''(x) = 2c_i + 6d_i(x - x_{i-1}) \quad \text{при} \quad x_{i-1} \leq x \leq x_i, \quad i = 1, \dots, n-1.$$

Получим еще $2(n-1)$ уравнений:

$$b_{i+1} = b_i + 2c_i h_i + 3d_i h_i^2, \quad (2.24)$$

$$c_{i+1} = c_i + 3d_i h_i. \quad (2.25)$$

Недостающие два уравнения можно получить из граничных условий, предполагая нулевую кривизну сплайна на концах отрезка. Приравнявая значения второй производной в точках x_0 и x_n нулю, получим два недостающих уравнения:

$$S''(x_0) = c_1 = 0, \quad (2.26)$$

$$S''(x_n) = c_n + 3d_n h_n = 0. \quad (2.27)$$

Заметим, что граничные условия (2.26) – (2.27) дают погрешность $O(h^2)$ вблизи границы. Поэтому вместо них можно рекомендовать условие непрерывности $S''(x)$ в двух внутренних (приграничных) узлах. Если существуют другие условия поведения функции в точках x_0 и x_n , то уравнения (2.26), (2.27) будут

другими. Таким образом, система (2.22) – (2.27) для отыскания $4n$ неизвестных $a_i, b_i, c_i, d_i, i = 1, 2, \dots, n$, замкнута. Её удобно решать, проведя предварительно следующие преобразования. Коэффициент a_i сразу получается из решения уравнения (2.22). Из уравнений (2.25) и (2.27) следует:

$$d_i = (c_{i+1} - c_i)/3h_i, \quad i = 1, 2, \dots, n, \quad (2.28)$$

$$d_n = -c_n/3h_n. \quad (2.28a)$$

Подставляя выражения (2.28), (2.28a) в (2.23) и заменяя в нем a_i на f_{i-1} , получаем

$$b_i = [(f_i - f_{i-1})/h_i] - \frac{1}{3}h_i(c_{i+1} + 2c_i), \quad (2.29)$$

$$b_n = [(f_n - f_{n-1})/h_n] - \frac{2}{3}h_n c_n. \quad (2.29a)$$

Из (2.24) и (2.25) имеем

$$b_{i+1} = b_i + h_i(c_i + c_{i+1}) \quad (2.29б)$$

или

$$b_i = b_{i-1} + h_{i-1}(c_{i-1} + c_i). \quad (2.29в)$$

Полагая в (2.29) $i = n$ и сравнивая с (2.29a), имеем

$$c_{n+1} = 0 \quad (2.29г)$$

Подставляя теперь (2.29) и (2.29a) в (2.29в), получаем систему уравнений относительно c_i , имеющую следующий вид:

$$c_1 = 0, \quad (2.30)$$

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = 3[(f_i - f_{i-1})/h_i - (f_{i-1} - f_{i-2})/h_{i-1}], \quad i = 2, \dots, n, \quad (2.30a)$$

$$c_n = 0. \quad (2.30б)$$

Матрица системы (2.30), (2.30a), (2.30б) – трехдиагональная, решение находится методом прогонки. Решив её, находим затем коэффициенты b_i из уравнений (2.29), (2.29a) и d_i – из уравнений (2.28), (2.28a).

Метод сплайн-интерполяции приводит к удовлетворительным результатам при интерполяции непрерывных функций с гладкими 1-й и 2-й производными. Причем кубический интерполяционный сплайн, построенный по узлам $f_i = f(x_i), i = 0, \dots, n$, будет обладать минимумом кривизны по сравнению с любой интерполяционной функцией, имеющей непрерывные первую и вторую производные. При сплайн-интерполяции функций с резким изменением производных сплайна могут иметь место большие ошибки. Сплаины более высокого порядка, чем третий, используются редко, так как при вычислении большого числа коэффициентов может накапливаться ошибка, приводящая к значительным отклонениям.

Рассмотрим еще сплайн второго порядка. Уравнение для него:

$$S(x) = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2,$$

$$x_{i-1} \leq x \leq x_i, \quad i = 1, 2, \dots, n.$$

Очевидно, что

$$S(x_{i-1}) = a_i = f_{i-1},$$

$$S(x_i) = a_i + b_i h_i + c_i h_i^2 = f_i, \quad (2.31)$$

$$h_i = x_i - x_{i-1}.$$

Условие совпадения первой производной $S'(x) = b_i + 2c_i(x - x_{i-1})$ в узлах дает:

$$b_i = f'_0 \text{ при } i = 0,$$

$$b_{i+1} = b_i + 2c_i h_i, \quad i = 1, 2, \dots, n-1. \quad (2.32)$$

Аналогично замечанию к (2.26) – (2.27) следует рекомендовать вместо (2.32) условие непрерывности $S''(x)$ в одном узле. Если принять, что в начальной точке задано f_0 и f'_0 , т.е. a_1 и b_1 , то для определения коэффициентов a_i, b_i, c_i на каждом участке нужно решать систему линейных уравнений третьего порядка (2.31), (2.32). Линейный сплайн или сплайн первого порядка представляет собой ломаную линию, состоящую из отрезков прямых, соединяющих узлы.

2.8. ПРИБЛИЖЕНИЕ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ (МНК)

До сих пор рассматривалась интерполяция, т.е. такой способ приближения, когда значения приближаемой и приближающей функций совпадают в узлах некоторой сетки. Однако достаточно часто, например, при аппроксимации большого числа экспериментальных точек, найденных с некоторой погрешностью, интерполяция становится неразумной. В этом случае целесообразно строить приближающую функцию таким образом, чтобы сгладить влияние погрешности измерения и числа точек эксперимента. Такое сглаживание реализуется при построении приближающей функции по методу наименьших квадратов. Вид приближающей функции может быть произвольным, ниже рассмотрен случай, когда приближающая функция является многочленом. При этом добиваются минимизации суммы квадратов отклонений значений приближаемой и приближающей функций в узлах сетки, называемой квадратичным отклонением.

Пусть задана таблично в узлах x_j функция y_j при $j = 0, 1, \dots, N$. Необходимо построить такой многочлен $F_n(x) = \sum_{i=0}^n a_i x^i$, для которого минимально квадратичное отклонение

$$\Phi = \sum_{j=0}^N [F_n(x_j) - y_j]^2. \quad (2.33)$$

Очевидно, что минимума Φ можно добиться только за счет изменения коэффициентов многочлена $F_n(x)$. Необходимые условия экстремума имеют вид

$$\frac{\partial \Phi}{\partial a_k} = 2 \sum_{j=0}^N \left[\sum_{i=0}^n a_i x_j^i - y_j \right] x_j^k = 0, \quad k = 0, 1, \dots, n. \quad (2.34)$$

Эту систему для удобства преобразуют к следующему виду:

$$\sum_{i=0}^n a_i \sum_{j=0}^N x_j^{k+i} = \sum_{j=0}^N y_j x_j^k, \quad k = 0, 1, \dots, n. \quad (2.35)$$

Система (2.35) называется нормальной системой метода наименьших квадратов и представляет собой систему линейных алгебраических уравнений относительно коэффициентов a_i . Решив систему, построим многочлен $F_n(x)$, приближающий функцию $f(x)$ и минимизирующий квадратичное отклонение.

Рассмотрим одно важное свойство системы (2.35). Предположим, что точки равномерно распределены на отрезке $[0, 1]$, т.е. $x_j = (1/N)j$, $j = 0, 1, \dots, N$.

В этом случае $\sum_{j=0}^N x_j^{k+i}$ можно приближенно заменить интегралом:

$$\sum_{j=0}^N x_j^{k+i} = N \sum_{j=0}^N \frac{1}{N} x_j^{k+i} = N \int_0^1 x^{k+i} dx = \frac{N}{k+i+1}.$$

Тогда определитель системы (2.35) будет

$$\Delta = N^{n+1} \det H_{n+1},$$

где H_{n+1} – матрица Гильберта порядка $n+1$ с элементами $[1/(k+i+1)]$.

Определитель этой матрицы

$$\det H_p = \frac{[1! 2! \dots (p-1)!]^3}{p!(p+1)! \dots (2p-1)!}$$

и очень быстро убывает с ростом p , что приводит к быстрому убыванию величины определителя системы Δ . Так, для

$$\begin{aligned} p = 2 (n = 1) \quad \det H_2 &\cong 10^{-2}, \\ p = 3 (n = 2) \quad \det H_3 &\cong 10^{-5}. \end{aligned}$$

Следовательно, система (2.35) с увеличением степени n приближающего многочлена становится плохо обусловленной и решение её связано с большой потерей точности. Поэтому при использовании метода наименьших квадратов, как правило, используют приближающий многочлен не выше третьей степени.

При необходимости построения многочленов большей степени пользуются приемами, позволяющими повысить обусловленность системы (2.35). Применяют обычно два приема. Первый – использование системы точек, позволяющей разбить систему (2.35) на две подсистемы меньшего порядка. Второй – использование системы ортогональных многочленов.

Как отмечалось выше, метод наименьших квадратов широко используется для сглаживания экспериментальных кривых, полученных с некоторой погрешностью. Если степень аппроксимирующего многочлена равна числу точек, то среднеквадратичный многочлен совпадает с интерполяционным. Поэтому хорошее сглаживание будет при $n \leq N$. Но если n очень мало, то для описания сложной кривой коэффициентов может не хватить. Для выбора оптимальной степени многочлена поступают следующим образом. Строят многочлен по методу наименьших квадратов некоторой степени n . Вычисляют квадратичное отклонение Φ и сравнивают его с известной величиной погрешности ε .

Если $\Phi \geq \varepsilon$, т.е. математическая ошибка существенно превышает ошибку экспериментальных данных, то степень приближающего многочлена недостаточна для описания кривой. Если же $\Phi \leq \varepsilon$, то старшие коэффициенты аппроксимации физически недостоверны. Хорошее сглаживание получается в том случае, если $\Phi \approx \varepsilon$ – степень приближающего многочлена оптимальна. Обычно начинают построение приближающего многочлена для случая $n=1$ и увеличивают его степень до тех пор, пока Φ не станет примерно равным ε . Если при этом $n \ll N$, то приближающий многочлен выбран верно. Если нет, то следует поискать более удачный вид приближающей функции.

Можно показать методами теории вероятностей, что функция $F_n(x)$, найденная по формулам (2.34), (2.35), с наибольшим правдоподобием совпадает с оценкой математического ожидания для заданной выборки $(f_0, f_1, \dots, f_n)$, т.е. метод наименьших квадратов обладает сглаживающими свойствами.

2.9. ПОСТАНОВКА ЗАДАЧИ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Задача численного дифференцирования возникает при нахождении производных от функции $y = f(x)$, заданной таблично, либо при нахождении производной от аналитической функции, непосредственное дифференцирование которой по каким-либо причинам затруднено. Важное приложение численного дифференцирования – разностная аппроксимация производных, которая широко используется при численном решении обыкновенных дифференциальных уравнений и дифференциальных уравнений в частных производных.

Общий подход к задаче численного дифференцирования состоит в замене функции $y = f(x)$ некоторой приближающей, легко вычисляемой функцией $\varphi(x, \bar{a})$, $y = \varphi(x, \bar{a}) + R(x)$, где $R(x)$ – остаточный член приближения. Полагают, что $y'(x) = \varphi'(x, \bar{a}) + R'(x)$, а $R'(x) = y'(x) - \varphi'(x, \bar{a})$ – погрешность вычисления производной.

Наиболее исследованным и широко распространенным является случай, когда в качестве приближающей функции $\varphi(x, \bar{a})$ берется интерполяционный многочлен. При этом производные соответствующих порядков легко определяются дифференцированием многочлена, а погрешность дифференцирования – дифференцированием остаточного члена или погрешности интерполяции, которая, согласно формуле (2.18), равна

$$R_n(x) = \varepsilon_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \omega_{n+1}(x).$$

2.10. ДИФФЕРЕНЦИРОВАНИЕ ИНТЕРПОЛЯЦИОННОГО МНОГОЧЛЕНА НЬЮТОНА

Для вывода формул численного дифференцирования удобно пользоваться интерполяционным многочленом в форме Ньютона, так как он содержит разделенные разности, являющиеся аналогом производных соответствующих порядков. Очевидно, что минимальное число узлов для получения k -й производной равно $(k+1)$, так как дальнейшее дифференцирование многочлена приводит к производной, равной нулю.

Запишем следующую формулу:

$$f(x) = P_n(x) + R_n(x). \quad (2.36)$$

Очевидно, что

$$f'(x) = P'_n(x) + R'_n(x), \quad (2.37')$$

$$f''(x) = P''_n(x) + R''_n(x), \quad (2.37'')$$

$$f^{(k)}(x) = P^{(k)}_n(x) + R^{(k)}_n(x). \quad (2.37''')$$

Для равноотстоящих узлов с шагом h имеем:

$$R_n^{(k)}(x) = O(h^{n+1-k}).$$

Таким образом, точность расчета производной при заданной степени интерполяционного многочлена уменьшается с увеличением номера производной. При решении практических задач, в том числе при численном решении обыкновенных дифференциальных уравнений и уравнений в частных производных, приходится использовать аппроксимации первых и вторых производных. Значительно реже приходится аппроксимировать производные более высоких порядков.

В связи с этим проведено вычисление первой и второй производной с использованием интерполяционных многочленов Ньютона первой и второй степени.

Первая производная. Из (2.15) и (2.16) с использованием (2.18), (2.36) и (2.37) получаем, дифференцируя по x :

$$f'(x) = P'_1(x) + R'_1(x) = \frac{f_0 - f_1}{x_0 - x_1} + [2x - (x_0 + x_1)] \frac{f''(\xi)}{2!}, \quad (2.38)$$

$$f'(x) = P'_2(x) + R'_2(x) = \frac{f_0 - f_1}{x_0 - x_1} + [2x - (x_0 + x_1)] \frac{\frac{f_0 - f_1}{x_0 - x_1} - \frac{f_1 - f_2}{x_1 - x_2}}{x_0 - x_2} + [(x - x_0)(x - x_1) + (x - x_1)(x - x_2) + (x - x_0)(x - x_2)] \frac{f'''(\xi)}{3!}. \quad (2.39)$$

Формулы (2.38) и (2.39) дают значения первой производной при неравноотстоящих узлах, при этом (2.38) получена дифференцированием интерполяционного многочлена первой степени, а (2.39) – второй степени с учетом формул для погрешностей интерполяции. Рассмотрим более подробно эти формулы. Очевидно, что при таблично заданных функциях с использованием интерполяционных многочленов производные вычисляются как производные от интерполяционных многочленов соответствующих порядков. Приближенное значение первой производной по формуле (2.38), полученное дифференцированием $P'_1(x)$, равно

$$f'(x) \approx P'_1(x) = \frac{f_0 - f_1}{x_0 - x_1} = \text{const}. \quad (2.40)$$

При вычислении по формуле (2.40) значения производной, которая, как видно, постоянна на всем интервале, максимальная ошибка, согласно (2.38), имеет место в крайних точках интервала x_0 и x_1 и пропорциональна первой степени $h = (x_0 - x_1)$. В средней точке $\bar{x}_1 = (x_0 + x_1)/2$, $R'_1(\bar{x}_1) = 0$, и значение производной вычисляется с более высокой точностью, чем в крайних узлах. Естественно, дело не в том, что в точке $\bar{x}_1$ получается точное значение производной, а в том, что выражение для $R'_n(x)$ является приближенным. Погрешность в точке x_0 :

$$R'_1(x_0) = -(x_1 - x_0) \frac{f''(\xi)}{2!},$$

в точке x_1 :

$$R'_1(x_1) = (x_1 - x_0) \frac{f''(\xi)}{2!},$$

т.е. равна по величине и противоположна по знаку погрешности в точке x_0 . Геометрическое представление формулы и аппроксимация производной показаны на рис. 2.4, а.

Видно, что угол наклона касательной в точке x_0 у функции $f(x)$ больше, а в точке x_1 меньше, чем угол наклона прямой $P'_1(x)$. В некоторой точке x^* точное значение производной и производная $P'_1(x)$ совпадают, и эта точка близка к точке $\frac{(x_0 + x_1)}{2}$.

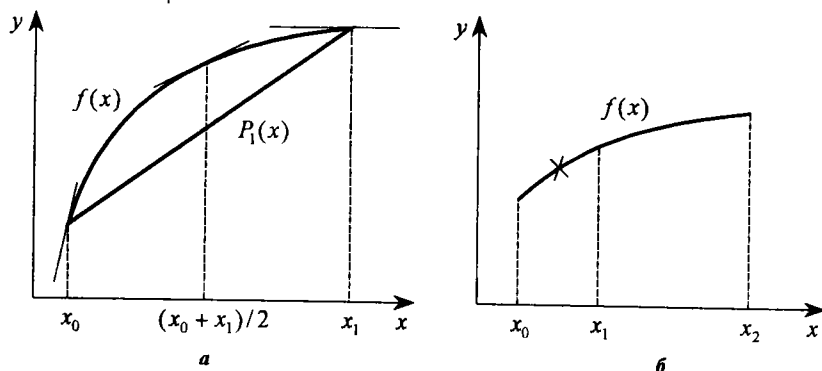


Рис. 2.4

При использовании интерполяционного многочлена второй степени имеем:

$$f'(x) \cong \frac{f_0 - f_1}{x_0 - x_1} + [2x - (x_0 + x_1)] \frac{\frac{f_0 - f_1}{x_0 - x_1} - \frac{f_1 - f_2}{x_1 - x_2}}{x_0 - x_2}. \quad (2.41)$$

Видно, что согласно (2.41) $f'(x)$ не есть константа на интервале $[x_0, x_2]$. Из формулы (2.41) и (2.39) следует, что в этом случае при почти равноотстоящих узлах максимальная ошибка в вычислении производной имеет место в крайних точках, а минимальная – в средней, при этом ошибка пропорциональна произведению двучлена, т.е. $(x_0 - x_1)(x_0 - x_2)$. Если узлы существенно неравноотстоящие, то ошибка может быть пропорциональна первой степени $(x - x_1)$. Из рис. 2.4, б очевидно, что если узел x_2 значительно удален и точка интерполяции расположена в интервале $[x_0, x_1]$, то погрешность вычисления производной пропорциональна первой степени шага $(x_0 - x_1)$, а не второй.

Вторая производная. Для вычисления второй производной, очевидно, нужно использовать интерполяционный многочлен Ньютона второй степени. Имеем:

$$f''(x) = P_2''(x) + R_2''(x) = \frac{2 \left[\frac{f_0 - f_1}{x_0 - x_1} - \frac{f_1 - f_2}{x_1 - x_2} \right]}{x_0 - x_2} + (3x - x_0 - x_1 - x_2) \frac{f''(\xi)}{3!}. \quad (2.42)$$

Таким образом, приближенное значение второй производной на интервале $[x_0, x_2]$ есть константа, равная

$$f''(x) \cong P_2''(x) + R_2''(x) = \frac{2 \left[\frac{f_0 - f_1}{x_0 - x_1} - \frac{f_1 - f_2}{x_1 - x_2} \right]}{x_0 - x_2}. \quad (2.42a)$$

Из (2.42) и (2.42a) следует, что по-прежнему максимальная погрешность в вычислении имеет место в крайних точках и формально по (2.42) равна нулю в точке $\bar{x}_2 = (x_0 + x_1 + x_2)/3$. Хотя в силу приближенности формулы (2.18) равна нулю в этой точке не строго. Отметим, что точки $x_k = (x_0 + x_1 + \dots + x_k)/(k+1)$ называются точками повышенной точности для производной k -го порядка.

2.11. БЕЗРАЗНОСТНЫЕ ФОРМУЛЫ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ ДЛЯ РАВНООТСТОЯЩИХ УЗЛОВ

Если принять, что узлы равноотстоящие и шаг сетки равен h , то из (2.38), (2.39) можно получить следующие формулы для расчета первой производной в узлах.

Из (2.38) имеем в точках x_0 и $x_1 = x_0 + h$:

$$f'(x_0) \cong \frac{f_0 - f_1}{x_0 - x_1} - h \frac{f''(\xi)}{2!}, \quad (2.43')$$

$$f'(x_1) \cong \frac{f_0 - f_1}{x_0 - x_1} + h \frac{f''(\xi)}{2!}. \quad (2.43'')$$

Из (2.39) имеем в точках x_0 , $x_1 = x_0 + h$, $x_2 = x_0 + 2h$:

$$f'(x_0) \cong \frac{1}{2h} [-3f_0 + 4f_1 - f_2] + \frac{h^2}{3} f'''(\xi), \quad (2.44')$$

$$f'(x_1) \cong \frac{1}{2h} [f_2 - f_0] - \frac{h^2}{6} f'''(\xi), \quad (2.44'')$$

$$f'(x_2) \cong \frac{1}{2h} [f_0 - 4f_1 + 3f_2] + \frac{h^2}{3} f'''(\xi). \quad (2.44''')$$

Из (2.42) имеем в точках x_0, x_1, x_2 следующие выражения для второй производной:

$$f''(x_0) \cong \frac{1}{h^2} [f_0 - 2f_1 + f_2] - h f'''(\xi), \quad (2.45')$$

$$f''(x_1) \cong \frac{1}{h^2} [f_0 - 2f_1 + f_2] - \frac{h^2}{12} f^{IV}(\xi), \quad (2.45'')$$

$$f''(x_2) \cong \frac{1}{h^2} [f_0 - 2f_1 + f_2] + h f'''(\xi). \quad (2.45''')$$

Говорят, что формулы численного дифференцирования (2.43'), (2.43''), (2.45'), (2.45''') – первого порядка точности (относительно h), а формулы (2.44') – (2.44''') и (2.45'') – второго порядка точности. Формулы (2.43) – (2.45) носят название безразностных формул численного дифференцирования. Видно, что при использовании формул (2.43) первая производная в узлах x_0, x_1

аппроксимируется с первым порядком точности, а при использовании формул (2.44) – со вторым порядком точности.

Вторая производная согласно (2.45) в крайних точках x_0 и x_2 аппроксимируется с первым порядком точности, а в центральной x_1 – со вторым. В центральных точках производные аппроксимируются более точно, чем в крайних, что достаточно очевидно (рис. 2.4) и следует также из приведенных формул. Аппроксимация производных в крайних точках используется при численном решении задачи Коши и краевых задач для обыкновенных дифференциальных уравнений и уравнений в частных производных. Для повышения точности аппроксимации производных теоретически следует увеличивать степень интерполяционного многочлена.

2.12. ПРИМЕНЕНИЕ РЯДА ТЕЙЛОРА ДЛЯ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

На практике формулы численного дифференцирования находят с использованием ряда Тейлора. Разлагая функцию $f(x)$ в ряд Тейлора по степеням h , где h достаточно мало, имеем:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + \frac{h^4}{4!} f^{IV}(x) + \dots, \quad (2.46')$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f'''(x) + \frac{h^4}{4!} f^{IV}(x) + \dots \quad (2.46'')$$

Из (2.46') и (2.46'') имеем

$$f'(x) \cong \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6} f'''(x), \quad (2.47')$$

$$f'(x) \cong \frac{f(x) - f(x-h)}{h} + \frac{h}{2!} f''(x). \quad (2.47'')$$

Вычитая из (2.46') (2.46''), получаем

$$f'(x) \cong \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6} f'''(x). \quad (2.48)$$

Очевидно, что формулы для первой производной (2.43') и (2.47'), (2.43'') и (2.47''), (2.44'') и (2.48) представляют собой попарно эквивалентную запись для первой производной. При сравнении этих формул нужно полагать в (2.47') $x = x_0$, а в (2.47'') и (2.48) $x = x_1$.

В теории разностных схем производную, вычисленную по формуле (2.47'), называют односторонней правой производной или производной вперед, вычисленную по формуле (2.47'') – односторонней левой производной или производной назад, а по формуле (2.48) – центральной или симметричной производной.

Складывая (2.46') и (2.46''), имеем для второй производной:

$$f''(x) \cong \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{h^2}{12} f^{IV}(x). \quad (2.49)$$

Очевидно, что формулы (2.45'') и (2.49) эквивалентны, если их в (2.49) полагать $x = x_1$.

2.13. ПОСТАНОВКА ЗАДАЧИ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ

В тех случаях, когда при вычислении определенного интеграла

$$F = \int_a^b f(x) dx \quad (2.50)$$

невозможно найти первообразную или она очень сложна для вычислений, то обращаются к формулам численного интегрирования. При этом подынтегральную функцию $f(x)$ заменяют такой приближающей, что она, во-первых, близка в каком-то смысле к $f(x)$:

$$f(x) \approx \varphi(x, \bar{a}),$$

а, во-вторых, интеграл от $\varphi(x, \bar{a})$ легко вычисляется. Наиболее употребителен случай, когда подынтегральную функцию заменяют интерполяционным многочленом Лагранжа:

$$f(x) = L_n(x) + R_n(x), \quad (2.51)$$

где $R_n(x)$ – остаточный член интерполяции. Подставляя (2.51) в (2.50), получаем

$$F = \int_a^b L_n(x) dx + R, \quad \text{где } R = \int_a^b R_n(x) dx$$

– остаточный член формулы численного интегрирования или её погрешность.

Заменяя $f(x)$ интерполяционными многочленами различной степени, получают формулы численного интегрирования различного порядка точности.

2.14. ФОРМУЛЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ: ПРЯМОУГОЛЬНИКОВ, ТРАПЕЦИЙ, ПАРАБОЛ (СИМПСОНА)

Заменяем подынтегральную функцию, входящую в (2.50), интерполяционным многочленом Лагранжа нулевой степени, проходящим через середину отрезка – точку $\bar{x} = (a+b)/2$ (рис. 2.5).

Из геометрических соображений очевидно, что площадь криволинейной трапеции можно заменить площадью прямоугольника, т.е.

$$F = \int_a^b f(x) dx \approx (b-a) f((a+b)/2). \quad (2.52)$$

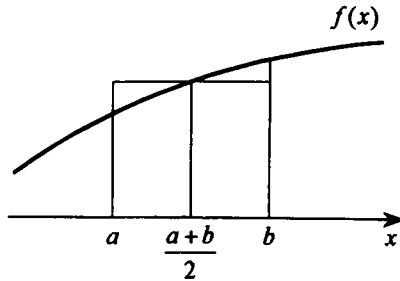


Рис. 2.5

Формула (2.52) и носит название формулы средних. Оценим её погрешность:

$$R = \int_a^b f(x) dx - (b-a)f(\bar{x}). \quad (2.53)$$

Используем разложение функции $f(x)$ в ряд относительно середины отрезка:

$$f(x) = f(\bar{x} - (x - \bar{x})) = f(\bar{x}) - (x - \bar{x})f'(\bar{x}) + \frac{(x - \bar{x})^2}{2!}f''(\bar{x}) + \dots \quad (2.54)$$

Подставив (2.54) в (2.53), получим

$$\begin{aligned} R &\approx \int_a^b \left[f(\bar{x}) - (x - \bar{x})f'(\bar{x}) + \frac{(x - \bar{x})^2}{2!}f''(\bar{x}) \right] dx - (b-a)f(\bar{x}) = f(\bar{x})(b-a) + \\ &+ \frac{1}{2}f''(\bar{x}) \left[\left(\frac{a-b}{2} \right)^2 - \left(\frac{b-a}{2} \right)^2 \right] + \frac{1}{6}f'''(\bar{x}) \left[\left(\frac{a-b}{2} \right)^3 - \left(\frac{b-a}{2} \right)^3 \right] - (b-a)f(\bar{x}) = \\ &= \frac{1}{24}f'''(\bar{x})(b-a)^3. \end{aligned} \quad (2.55)$$

При вычислении ошибки интегрирования уничтожился не только первый, но и второй член разложения, что связано с симметричным выбором узла интегрирования. И хотя по построению формула точна для многочленов нулевого порядка, выбор симметричного узла интерполяции привел к тому, что формула точна для любой линейной функции.

Величина остаточного члена формулы прямоугольников может быть достаточно велика, так как величина отрезка $(b-a)$ может быть достаточно большой. Для повышения точности введем сетку

$$a = x_0 \leq x_1 \leq \dots \leq x_N = b$$

с достаточно мелким шагом $h_i = x_i - x_{i-1}$ и применим формулу прямоугольников на каждом шаге сетки. Тогда получим обобщенную формулу прямоугольников

$$\int_a^b f(x) dx \approx \sum_{i=1}^N h_i f\left(\frac{x_{i-1} + x_i}{2}\right) \quad (2.56)$$

с величиной остаточного члена

$$R \approx \frac{1}{24} \sum_{i=1}^N h_i^3 f''\left(\frac{x_{i-1} + x_i}{2}\right).$$

На равномерной сетке с шагом $h_i = x_i - x_{i-1} = \text{const}$ формула (2.56) упрощается и имеет вид

$$\int_a^b f(x) dx \approx h \sum_{i=1}^N f\left(\frac{x_{i-1} + x_i}{2}\right), \quad (2.57)$$

величина остаточного члена

$$R \approx \frac{1}{24} h^3 \sum_{i=1}^N f''\left(\frac{x_{i-1} + x_i}{2}\right). \quad (2.58)$$

Заменяя в (2.58) сумму интегралом, получаем

$$R \approx \frac{1}{24} h^2 \int_a^b f''(x) dx. \quad (2.59)$$

Для справедливости оценки (2.58) необходимо существование непрерывной второй производной; если $f''(x)$ – кусочно-непрерывная, то удастся сделать лишь мажорантную оценку, заменяя $f''(x)$ её максимальной величиной на $[a, b]$:

$$R \leq \frac{1}{24} h^2 M_2 (b-a), \quad (2.60)$$

где $M_2 = \max |f''(x)|_{[a,b]}$.

В том случае, если функция $f(x)$ задана в виде таблицы, её значение в середине интервала неизвестно. Это значение находится, как правило, интерполированием, что приводит к ухудшению точности формулы.

В случае таблично заданных функций удобно в качестве узлов интерполяции выбрать начало и конец отрезка интегрирования, т.е. заменить функцию $f(x)$ многочленом Лагранжа первой степени. Имеем

$$\int_a^b f(x) dx = \int_a^b L_1(x) dx + \int_a^b R_1(x) dx.$$

В этом случае величина интеграла, равная площади криволинейной трапеции, приближенно заменяется величиной площади трапеции (рис. 2.6).

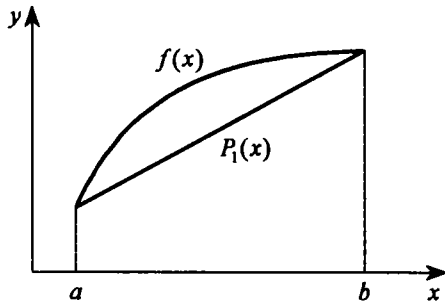


Рис. 2.6

Поэтому

$$\int_a^b f(x) dx \approx \frac{1}{2}(b-a)[f(a) + f(b)] = \int_a^b L_1(x) dx, \quad (2.61)$$

имея в виду, что $x_0 = a$, $x_1 = b$. Эта формула носит название формулы трапеций.

Для оценки погрешности интегрирования при использовании формулы трапеций вычислим $\int_a^b R_1(x) dx$, используя формулы (2.18). Имеем

$$R = \int_a^b R_1(x) dx = -\frac{1}{12}(b-a)^3 f''(\xi). \quad (2.62)$$

Величина погрешности формулы трапеций вдвое больше, чем формулы прямоугольников. Это объясняется тем, что выбор в формуле прямоугольников в качестве узла интерполяции симметричного узла приводит к повышению её точности.

Для повышения точности формулы (2.61) введем на $[a, b]$ сетку

$$a = x_0 \leq x_1 \leq \dots \leq x_N = b.$$

Подсчитывая значение интеграла для каждого интервала и суммируя эти значения, получаем обобщенную формулу трапеций:

$$F(x) = \int_a^b f(x) dx \approx \frac{1}{2} \sum_{i=1}^N (f_{i+1} + f_i)(x_{i+1} - x_i), \quad (2.63)$$

а величина остаточного члена

$$R \approx -\frac{1}{12} \sum_{i=1}^N (x_i - x_{i-1})^3 f''\left(\frac{x_{i+1} + x_i}{2}\right).$$

На сетке с постоянным шагом $h = h_i = x_i - x_{i-1} = \text{const}$, $i = 0, 1, \dots, N-1$ эти формулы упрощаются:

$$F = \int_a^b f(x) dx \approx h \left[\frac{1}{2}(f_0 + f_N) + \sum_{i=1}^{N-1} f_i \right], \quad (2.64)$$

$$R \approx -\frac{1}{12} h^3 \sum_{i=1}^N f''\left(\frac{x_i + x_{i-1}}{2}\right) \approx -\frac{1}{12} h^2 \int_a^b f''(x) dx. \quad (2.65)$$

На практике пользуются мажорантной оценкой величины остаточного члена

$$|R| \leq \frac{b-a}{12} h^2 M_2, \quad (2.66)$$

где $M_2 = \max_{[a,b]} |f''(x)|$.

Таким образом, формула трапеций (как и формула прямоугольников) имеет второй порядок точности относительно шага сетки, и погрешность асимптотически стремится к нулю при $h \rightarrow 0$ с точностью до членов более высокого порядка малости.

Для повышения порядка точности формулы численного интегрирования заменим подынтегральную кривую параболой – интерполяционным многочленом Лагранжа второй степени, выбрав в качестве узлов интерполяции концы и середину отрезка интегрирования: $x_0 = a$, $x_1 = (a+b)/2$, $x_2 = b$ (рис 2.7).

В этом случае, проинтегрировав интерполяционный многочлен для равноотстоящих узлов, получим

$$F \approx \int_a^b L_2(x) dx = \frac{h}{3} [f_0 + 4f_1 + f_2], \quad (2.67)$$

а величина остаточного члена $R = \int_a^b R_2(x) dx$,

$$R \approx -\frac{(b-a)^5}{180} f^{IV}\left(\frac{a+b}{2}\right).$$

Формулу (2.67) называют формулой Симпсона. Для неравноотстоящих узлов x_0, x_1, x_2 $F \approx \int_a^b L_2(x) dx$.

Как и в предыдущих двух случаях, для повышения точности формулы (2.67) введем сетку с достаточно малым шагом. Суммируя значения интегралов, полученных по (2.67) для каждого интервала, получаем обобщенную формулу Симпсона (парабол), которая на равномерной сетке имеет вид

$$F \approx \frac{h}{3} [f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + \dots + 2f_{N-2} + 4f_{N-1} + f_N], \quad (2.68)$$

а величина остаточного члена

$$R \approx -\frac{h^4}{180} \int_a^b f^{IV}(x) dx, \quad (2.69)$$

т.е. формула парабол имеет четвертый порядок точности относительно шага сетки. Как правило, для оценки величины погрешности пользуются мажорантной оценкой

$$|R| \leq \frac{(b-a)}{180} h^4 M_4, \quad (2.70)$$

где $M_4 = \max |f^{IV}(x)|_{[a,b]}$.

Количество интервалов N , на которое делится отрезок интегрирования, должно быть всегда четным, так как интерполяционный многочлен второй степени проводится через три точки.

2.15. МЕТОД РУНГЕ-РОМБЕРГА-РИЧАРДСОНА ПОВЫШЕНИЯ ПОРЯДКОВ ТОЧНОСТИ

При вычислении определенного интеграла на сетке с большим числом узлов получается более точный результат, но эти формулы более громоздки. Метод Рунге-Ромберга-Ричардсона позволяет получать более высокий порядок точности без значительного увеличения числа операций.

Пусть для вычисления величины $f(x)$ имеется некоторая формула $\varphi(x, h)$, дающая возможность приближенного расчета величины $f(x)$ на равномерной сетке с шагом h . Величину остаточного члена можно представить в следующем виде:

$$f(x) - \varphi(x, h) = \psi(x)h^p + O(h^{p+1}), \quad (2.71)$$

где p – порядок точности расчетной формулы, а $(x)h^p$ – главный член погрешности. Например, для формулы прямоугольников и трапеций $p=2$, а для формулы парабол $p=4$.

Проведя расчет на другой сетке с шагом rh и получив другое приближенное значение, $\varphi(x, rh)$ величины $f(x)$, будем иметь

$$f(x) - \varphi(x, rh) = \psi(x)(rh)^p + O((rh)^{p+1}). \quad (2.72)$$

Вычитая из (2.71) (2.72), получаем

$$(x)h^p = \frac{\varphi(x, h) - \varphi(x, rh)}{r^p - 1} + O(h^{p+1}). \quad (2.73)$$

Следовательно, пользуясь расчетом на сетке с шагом rh , удастся оценить главный член погрешности расчета на первой сетке. Если же подставить найденную погрешность (2.73) в (2.72), то получим результат с более высокой точностью:

$$f(x) = \varphi(x, h) + \frac{\varphi(x, h) - \varphi(x, rh)}{r^p - 1} + O(h^{p+1}). \quad (2.74)$$

Метод Рунге-Ромберга-Ричардсона может быть использован для оценки погрешности расчета и получения результата с более высокой степенью точности при решении различных задач, когда возможен расчет какой-либо величины на сетках с различным шагом, например, при численном интегрировании, дифференцировании, решении дифференциальных уравнений.

В качестве примера рассмотрим полученные формулы численного интегрирования, имеющие более высокий порядок точности. Проведем расчет определенного интеграла по формуле трапеций с шагом $2h$, (точки x_0, x_2), а затем на сетке с шагом h , (точки x_0, x_1, x_2). Обозначим результат этих расчетов F_{2h} и F_h . Порядок точности формулы трапеций равен двум, а сам остаточный член имеет такой же вид, как и в формуле (2.71). Обозначая более точное значение F , проведем уточнение по формуле (2.74):

$$\begin{aligned} F &= F_{2h} + \frac{F_{2h} - F_h}{(1/2)^2 - 1} = \frac{1}{3} [4F_h - F_{2h}] = \frac{1}{3} \left[4 \frac{h}{2} [f_0 + 2f_1 + f_2] - \frac{2h}{2} [f_0 + f_2] \right] = \\ &= \frac{h}{3} [f_0 + 4f_1 + f_2]. \end{aligned}$$

Таким образом, уточнение расчета формулы трапеций с использованием (2.74) привело к формуле парабол, имеющей более высокий – четвертый – порядок точности.

Глава 3. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

3.1. КРАТКОЕ СОДЕРЖАНИЕ ГЛАВЫ

В главе рассматриваются различные методы решения задачи Коши и краевой задачи для обыкновенных дифференциальных уравнений. Подробно рассмотрен метод Эйлера и ошибки, возникающие при реализации метода на ЭВМ; семейство одношаговых методов Рунге-Кутты; многошаговые методы. Из методов решения краевой задачи рассматриваются метод стрельбы и разностный метод.

3.2. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ (ОДУ)

1. Обыкновенным дифференциальным уравнением называется уравнение вида

$$F(x, y, y', \dots, y^{(n)}) = 0, \quad (3.1)$$

связывающее одну независимую переменную x , искомую функцию $y(x)$ и ее производные вплоть до n -го порядка.

2. Порядком ОДУ называется порядок старшей производной от искомой функции (например, $y'' - x^2y + x^2 = 0$ – ОДУ второго порядка).

3. ОДУ называется линейным, если оно имеет следующий вид:

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_1(x)y' + a_0(x)y + f(x) = 0.$$

Например, ОДУ $y'' - x^2y + x^2 = 0$ – линейное ОДУ второго порядка, а $y'' + e^y = 0$ – нелинейное.

4. Общим интегралом уравнения (3.1) называют функцию

$$\Phi(x, y, c_1, \dots, c_n), \quad (3.2)$$

связывающую независимую переменную x , искомую функцию $y(x)$ и n констант интегрирования с помощью уравнения $\Phi(x, y, c_1, \dots, c_n) = 0$, т.е. $y(x)$ входит в (3.2) неявным образом, причем количество констант интегрирования равно порядку ОДУ.

5. Общим решением ОДУ называется функция

$$y(x) = \varphi(x, c_1, \dots, c_n), \quad (3.3)$$

связывающая независимую переменную x и констант интегрирования, т.е. $y(x)$ определяется явным образом.

6. Для определения констант интегрирования $c_1, \dots, c_n$ задаются дополнительные условия в количестве, равном количеству констант интегрирования или порядку ОДУ.

Если в дополнительные условия подставить (3.2) и решить полученную систему относительно $c_1, \dots, c_n$, а затем последние подставить в (3.2), то получим частный интеграл $\Phi_1(x, y(x)) = 0$.

Аналогичные процедуры с общим решением (3.3) дают частное решение

$$y(x) = \varphi_1(x).$$

7. Если все дополнительные условия задаются в одной точке x_0 , то совокупность ОДУ и дополнительных условий называют задачей Коши для рассматриваемого ОДУ. В этом случае дополнительные условия называют начальными условиями.

8. Если дополнительные условия задаются более чем в одной граничной точке расчетной области, то совокупность ОДУ и дополнительных условий называют краевой задачей для рассматриваемого ОДУ, а дополнительные условия – граничными или краевыми условиями.

9. Таким образом, задача Коши для ОДУ (3.1) имеет вид

$$\begin{cases} F(x, y, y', \dots, y^{(n)}) = 0, \\ y(x_0) = y_0, \\ y'(x_0) = y_1, \\ \dots \\ y^{(n-1)}(x_0) = y_{n-1}, \end{cases}$$

где $y_0, y_1, \dots, y_{n-1}$ – заданные числа.

Таким образом, порядок старшей производной в начальных условиях не превышает $(n-1)$, где n – порядок ОДУ.

10. Если старшая производная в ОДУ выражается как функция x, y и производных более низкого порядка, то уравнение называют разрешенным относительно старшей производной и оно записывается в следующем виде: $y^{(n)} = f(x, y, y', \dots, y^{(n-1)})$.

11. Задачу Коши для ОДУ n -го порядка $y^{(n)} = f(x, y, y', \dots, y^{(n-1)})$ можно свести к задаче Коши для следующей нормальной системы ОДУ:

$$\begin{cases} y' = z_1(x), \\ z_1' = z_2(x), \\ \dots \\ z_{n-2}' = z_{n-1}(x), \\ z_{n-1}' = f(x, y(x), z_1(x), \dots, z_{n-1}(x)), \\ y(x_0) = y_0, \\ z_1(x_0) = y_1, \\ z_2(x_0) = y_2, \\ \dots \\ z_{n-1}(x_0) = y_{n-1}, \end{cases}$$

вводя обозначения:

$y'(x) = z_1(x)$ – первое дифференциальное уравнение,

$y''(x) = z_1'(x) = z_2(x)$ – второе дифференциальное уравнение,

...

$y^{(n-1)}(x) = \dots = z_{n-2}'(x) = z_{n-1}(x)$ – $(n-1)$ -е дифференциальное уравнение.

Важность задачи Коши во многих областях науки и техники явилась причиной разработки для ее решения большого количества методов. Ниже рассматриваются следующие две группы численных методов решения задачи Коши:

1. *Одношаговые методы*, в которых для нахождения решения в некоторой точке отрезка используется информация лишь в одной предыдущей точке. К этим методам относятся метод Эйлера, методы Рунге-Кутты.

2. *Многошаговые методы*, в которых для отыскания решения в некоторой точке используется информация о решении в нескольких предыдущих точках. К таким методам относится метод Адамса.

Ниже рассматривается применение этих методов к дифференциальным уравнениям первого порядка. Алгоритм для случая системы дифференциальных уравнений легко получается из алгоритма для решения одного уравнения.

Необходимо отметить, что эти методы можно применять только к хорошо обусловленным задачам, т.е. таким задачам, решение которых мало изменяется в случае малого изменения начальных условий. В противном случае небольшие ошибки, которые неизбежны при решении задачи, могут сильно изменить решение.

В качестве примера плохо обусловленной задачи рассмотрим решение следующей задачи Коши на $[0, 100]$:

$$u'(x) = u - x,$$

$$u(0) = 1.$$

Общее решение этого дифференциального уравнения такое:

$$u(x, c) = 1 + x + ce^x.$$

Из начального условия находим, что константа $c=0$, поэтому $u(100)=101$

Предположим, что начальное условие изменилось и теперь $u(0)=1.000001$.

Следовательно, изменится и константа c в общем решении. Теперь $c=10^{-6}$ и решение $u(100) \approx 2.7 \cdot 10^{37}$, т.е. изменение начального условия на одну миллионную изменило решение на 35 порядков.

Постановка краевых задач для ОДУ. Постановку краевых задач для ОДУ рассмотрим на примере ОДУ второго порядка с переменными коэффициентами:

$$y'' + p(x)y' + q(x)y = f(x), \text{ где } x \in [0, l].$$

1. Если на границах $x=0$ и $x=1$ заданы значения искомой функции, то такие граничные условия называются *граничными условиями первого рода*, а задачу

$$\begin{cases} y'' + p(x)y' + q(x)y = f(x), \\ y(0) = y_0, \\ y(l) = y_1 \end{cases}$$

называют *первой краевой задачей*, где y_0, y_1 – заданные числа.

2. Если на границах $x=0$ и $x=1$ заданы значения производных искомой функции, то такие граничные условия называются *граничными условиями второго рода*:

$$y'(0) = y_0,$$

$$y'(l) = y_1,$$

а задача

$$\begin{cases} y'' + p(x)y' + q(x)y = f(x), \\ y'(0) = y_0, \\ y'(l) = y_1 \end{cases}$$

– *второй краевой задачей*.

3. Если на границах задана линейная комбинация первой производной и самой искомой функции, то такие граничные условия называются *граничными условиями третьего рода*:

$$\alpha_0 y'(0) + \beta_0 y(0) = y_0,$$

$$\alpha_1 y'(l) + \beta_1 y(l) = y_1,$$

а задача

$$\begin{cases} y'' + p(x)y' + q(x)y = f(x), \\ \alpha_0 y'(0) + \beta_0 y(0) = y_0, \\ \alpha_1 y'(l) + \beta_1 y(l) = y_1 \end{cases}$$

– *третьей краевой задачей*.

Видно, что порядок старшей производной в краевых условиях не может превышать порядка ОДУ минус единица.

В заключение приведем одну из постановок краевой задачи для ОДУ четвертого порядка, встречающегося в теории упругости:

$$\begin{cases} y^{(4)} + p(x)y = f(x), \\ y(0) = y_0, \\ y'(0) = y_1, \\ y(l) = y_2, \\ y'(l) = y_3, \end{cases}$$

где y_0, y_1, y_2, y_3 – заданные числа.

3.3. ПОНЯТИЕ О МЕТОДЕ КОНЕЧНЫХ РАЗНОСТЕЙ. ПОРЯДОК ТОЧНОСТИ РАЗНОСТНОЙ СХЕМЫ

Численное решение обыкновенных дифференциальных уравнений и уравнений в частных производных во многих случаях осуществляется методом конечных разностей. Метод конечных разностей сводит решение дифференциальных уравнений к решению линейных или нелинейных уравнений с достаточно разреженными матрицами. При этом построение решения в методе сеток осуществляется в три этапа:

1. Область непрерывного изменения аргумента (или аргументов) заменяется конечным дискретным множеством точек, называемых разностной сеткой. В разностной сетке выделяются внутренние и граничные узлы. Решение разыскивается во внутренних узлах, а в граничных узлах значение искомой функции задается при аппроксимации граничных условий исходной дифференциальной задачи. Функция дискретного аргумента, определенная на разностной сетке, называется сеточной функцией.

2. Дифференциальные уравнения и граничные условия заменяются по определенным правилам своими разностными аналогами. Разностные операторы, соответствующие дифференциальному уравнению, записываются во внутренних узлах сетки. Разностные операторы, соответствующие граничным условиям, записываются в граничных узлах. В результате получается система алгебраических уравнений, число которых пропорционально числу внутренних узлов разностной сетки.

3. Осуществляется решение системы алгебраических уравнений каким-либо из известных методов. В большинстве случаев получаемая система уравнений есть система линейных алгебраических уравнений очень большого порядка (как правило, $N \geq 100$), но с очень разреженной матрицей.

В случае нелинейных систем итерационные процедуры, как правило, сводят их к линейным системам.

Продemonстрируем эти этапы на примере решения задачи Коши для простейшего линейного уравнения первого порядка. Пусть требуется найти решение задачи Коши на интервале $(0,1)$ для уравнения

$$\begin{aligned} \frac{du}{dx} + Au &= 0, \\ u(0) &= u_0. \end{aligned} \quad (3.4)$$

Здесь и в дальнейшем точное решение будем обозначать большой буквой U , а численное – малой u . Заменяем непрерывный интервал множеством точек $0, x_1, x_2, \dots, x_{n-1}, x_n, \dots, 1$.

Расстояние между точками называется шагом разностной сетки. В общем случае эти шаги могут быть переменными. Ниже рассмотрим случай постоянного шага и примем $x_n = nh$. Заменяем, далее, дифференциальное уравнение разностным в точке x_n :

$$\frac{du}{dx} + Au \approx \frac{u(x_{n-1} + h) - u(x_{n-1})}{h} + Au(x_{n-1}) = 0. \quad (3.5)$$

Таким образом, первая производная аппроксимирована односторонней разностью вперед с первым порядком точности. Перепишем уравнение (3.5) в виде

$$\frac{u_n - u_{n-1}}{h} + Au_{n-1} = 0. \quad (3.6)$$

Здесь $u(x_{n-1} + h) = u_n$, а $u(x_{n-1}) = u_{n-1}$.

Из (3.6) следует, что

$$u_n = (1 - Ah)u_{n-1}. \quad (3.7)$$

Из рекуррентного соотношения (3.7) имеем точное решение разностного уравнения (3.6):

$$u_n = u_0(1 - Ah)^n = u_0(1 - Ah)^{\frac{x_n}{h}}. \quad (3.8)$$

Очевидно, что точное решение разностных уравнений можно получить лишь для простейших случаев. Точное решение задачи (3.4) имеет вид

$$U(x_n) = u_0 e^{-Ax_n}. \quad (3.9)$$

Определим погрешность разностного решения. Имеем

$$\delta(x_n) = [e^{-Ax_n} - (1 - Ah)^{\frac{x_n}{h}}]u_0. \quad (3.10)$$

Предполагая малым шаг разностной схемы, представим $(1 - Ah)^{\frac{x_n}{h}}$ в виде

$$\begin{aligned}(1 - Ah)^{\frac{x_n}{h}} &= e^{\frac{x_n}{h} \ln(1 - Ah)} = e^{\frac{x_n}{h} \left[-Ah - \frac{A^2 h^2}{2} + O(h^3) \right]} = e^{-Ax_n} e^{-\frac{A^2 h}{2} x_n} e^{O(h^2)} = \\ &= e^{-Ax_n} \left[1 - \frac{A^2 h x_n}{2} + O(h^2) \right] [1 + O(h^2)] = e^{-Ax_n} - h \frac{A^2 x_n}{2} e^{-Ax_n} + O(h^2).\end{aligned}$$

Подставляя это выражение в (3.10), имеем

$$\delta(x_n) = hu_0 \frac{A^2 x_n}{2} e^{-Ax_n} + O(h^2) = O(h). \quad (3.11)$$

Таким образом, величина погрешности имеет первый порядок точности и совпадает в данном случае с порядком аппроксимации производной. В общем случае порядок погрешности решения определяется не только порядком аппроксимации дифференциального уравнения, но и порядком аппроксимации граничных условий.

Говорят, что разностная схема (3.5) имеет первый порядок точности на интервале. Можно показать, что на шаге эта разностная схема имеет второй порядок точности. Представим точное решение в виде ряда Тейлора в точке n . Имеем

$$U_n = u_{n-1} + h \left. \frac{du}{dx} \right|_n + \frac{h^2}{2!} \left. \frac{d^2 u}{dx^2} \right|_n + O(h^3). \quad (3.12)$$

Поскольку $\left. \frac{du}{dx} \right|_n = -Au_{n-1}$, а $\left. \frac{d^2 u}{dx^2} \right|_n = A^2 u_{n-1}$, то

$$U_n = (1 - Ah)u_{n-1} + A^2 u_{n-1} \frac{h^2}{2} + O(h^3). \quad (3.13)$$

Из (3.7) и (3.13) видно, что погрешность на шаге имеет порядок $O(h^3)$. Очевидно, что погрешность на интервале выше на порядок, так как она накапливается по мере увеличения числа шагов. Отметим еще, что погрешность решения стремится к нулю при $h \rightarrow 0$, т.е. решение разностной задачи сходится к точному. Однако для сходимости необходимы не только аппроксимация дифференциального уравнения, но и устойчивость разностной схемы. Не углубляясь в проблему устойчивости, отметим лишь, что свойство устойчивости можно трактовать как равномерную зависимость от h решения разностной задачи относительно возмущений правых частей и граничных условий.

В теории разностных схем доказывается, что разностная схема является сходящейся, если она обладает свойством аппроксимации и устойчива.

3.4. МЕТОД ЭЙЛЕРА. МЕТОД ЭЙЛЕРА С ПЕРЕСЧЕТОМ

Метод Эйлера является простейшим методом решения задачи Коши и имеет невысокую точность, поэтому на практике его используют достаточно редко. Однако на его основе в дальнейшем легче объяснить алгоритмы более эффективных (и, как правило, более сложных) методов и способы построения и исследования этих алгоритмов.

Рассмотрим задачу Коши для дифференциального уравнения первого порядка:

$$\begin{aligned}\frac{dU}{dx} &= f(x, U), \quad x \in [a, b], \\ U(a) &= U_0.\end{aligned} \quad (3.14)$$

Введем на $[a, b]$ сетку (в общем случае неравномерную) значений аргумента x такую, чтобы выполнялось следующее соотношение: $a = x_0 < x_1 < \dots < x_N = b$. Разложим решение $U(x)$ в окрестности узла сетки x_n по формуле Тейлора. Обозначая $U(x_n) = U_n$, имеем

$$U_{n+1} = U_n + h_n U'_n + \frac{1}{2} h_n^2 U''_n + \frac{1}{3!} h_n^3 U'''_n + \dots, \quad h_n = x_{n+1} - x_n. \quad (3.15)$$

Если функция f имеет непрерывную p -ю производную, то в (3.15) можно оставлять члены вплоть до $O(h^{p+1})$. Эти производные можно найти, дифференцируя правую часть (3.14) требуемое число раз. Например, для первой и второй производных имеем

$$\begin{aligned}U''(x) &= f(x, U), \\ U''(x) &= \frac{d}{dx} U'(x) = \frac{d}{dx} f(x, U) = f'_x U' + f'_U = ff'_x + f'_x = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial U} \frac{\partial U}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial U} f.\end{aligned}$$

Аналогично можно получить производные более высоких порядков.

Однако использование формулы (3.15) с большим числом членов имеет ряд недостатков: во-первых, с ростом порядка производной выражение для нее может оказаться очень сложным; кроме того, если функция f известна лишь приближенно или задана таблично, ее производные находятся с большой ошибкой.

Поэтому в разложении (3.15) оставляют только два члена. При такой замене вместо точного решения $U(x_{n+1})$ получается его приближенное значение $u(x_{n+1})$, которое находится по следующей формуле:

$$u_{n+1} = u_n + h_n f(x_n, u_n) \quad (3.16)$$

Так как значение $U(x_0) = U_0$ известно из начального условия, то последовательно пользуясь формулой (3.16), находим приближенные решения

$u_1, u_2, \dots, u_N$. Формула (3.16) записана для случая неравномерной сетки. Полагая шаг сетки постоянным, т.е. $x_{n+1} - x_n = h = \text{const}$, получим $u_{n+1} = u_n + h f(x_n, u_n)$.

Формула (3.16) $u_{n+1} = u_n + h f(x_n, u_n)$ является основной формулой метода Эйлера, или метода ломаных. Последнее название становится понятным из геометрической интерпретации схемы, представленной на рис. 3.1, на которой изображено семейство интегральных кривых уравнения (3.14).

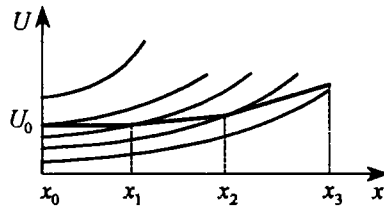


Рис. 3.1

Как видно из рисунка, решение задачи Коши по методу Эйлера дает решение, не совпадающее ни с одной из интегральных кривых, и является ломаной линией, совпадающей на каждом шаге с касательной к соответствующей интегральной кривой. Из рис. 3.1 следует, что метод Эйлера дает одностороннее приближение к точному решению $U(x)$. Из

сравнения формул (3.15) и (3.16) видно, что метод Эйлера обладает вторым порядком точности на шаге, являясь в то же время методом первого порядка аппроксимации на интервале.

Существует несколько модификаций метода Эйлера. Остановимся на одной из них – методе Эйлера с пересчетом, который называют также методом Эйлера-Коши. В этом методе значение u_{n+1} находится по формуле

$$u_{n+1} = u_n + \frac{h}{2} [f(x_n, u_n) + f(x_{n+1}, u_{n+1})], \quad (3.17)$$

т.е. вместо тангенса угла наклона касательной к интегральной кривой в точке x_n, u_n , который используется в методе Эйлера (см. (3.16)), используется полусумма значений тангенсов углов наклона касательных в известной (x_n, u_n) и искомой точке (x_{n+1}, u_{n+1}) . Поскольку, однако, u_{n+1} неизвестно, то (3.17) есть в общем случае нелинейное уравнение относительно u_{n+1} , которое можно решить различными методами (см. гл. 1).

В рассматриваемом случае логично использовать метод простой итерации (см. (1.10)), поскольку нелинейное уравнение уже разрешено относительно u_{n+1} . Тогда итерационный процесс запишется в виде

$$u_{n+1}^{(k+1)} = u_n + \frac{h}{2} [f(x_n, u_n) + f(x_{n+1}, u_{n+1}^{(k)})], \quad (3.18)$$

где верхний индекс означает номер итерации.

В качестве $u_{n+1}^{(0)}$ можно принять либо u_n , либо $u_{n+1}^{(0)} = u_n + h f(x_n, u_n)$, т.е. использовать значение, вычисленное по формуле Эйлера (3.17).

Тогда в первой итерации имеем

$$u_{n+1}^{(1)} = u_n + \frac{h}{2} \{f(x_n, u_n) + f[x_{n+1}, u_n + h f(x_n, u_n)]\}. \quad (3.19)$$

Формула (3.19) и есть основная формула метода Эйлера с пересчетом. Подобные схемы часто называют схемами типа «прогноз-корректор» («предиктор-корректор»). Сначала по формуле (3.16) получаем приближение (прогнозируемое) решения, а затем по формуле (3.19) это решение уточняется, т.е. корректируется. Нетрудно видеть, что метод Эйлера с пересчетом обладает третьим порядком точности на шаге. Действительно, из (3.15) имеем, что

$$U_{n+1} = U_n + h f(x_n, u_n) + \frac{h^2}{2} \left[\frac{\partial f}{\partial x} + f(x_n, u_n) \frac{\partial f}{\partial u} \right] + O(h^3). \quad (3.20)$$

Очевидно, что разлагая в ряд второй член в квадратных скобках (3.19), имеем

$$f[x_n + h, u_n + h f(x_n, u_n)] = f(x_n, u_n) + \frac{\partial f}{\partial x} h + \frac{\partial f}{\partial u} h f(x_n, u_n). \quad (3.21)$$

Подставляя (3.21) в (3.19), получаем

$$u_{n+1} = u_n + h f(x_n, u_n) + \frac{h^2}{2} \left[\frac{\partial f}{\partial x} + f(x_n, u_n) \frac{\partial f}{\partial u} \right]. \quad (3.22)$$

Сравнивая (3.20) и (3.22), видим, что на шаге метод Эйлера с пересчетом обладает третьим порядком точности и, соответственно, вторым на интервале. Метод Эйлера с пересчетом дает двустороннее приближение к решению.

Отметим, что поскольку нулевое приближение итерационного процесса задано, то итерационный процесс (3.18) можно продолжать до достижения заданной точности. Однако последующие итерации не повышают порядка точности схемы, и поэтому на практике используют 1 – 2 итерации. Ниже показано, что метод Эйлера и метод Эйлера с пересчетом являются частными случаями двухстадийных схем Рунге-Кутты.

3.5. МЕТОД РУНГЕ-КУТТЫ

Метод Рунге-Кутты позволяет строить схемы различного порядка точности. Основная идея метода состоит в построении специального алгоритма, такого, чтобы приращение функции на шаге $\Delta u = u_{n+1} - u_n$ совпадало с приращением Δu , которое определяется из ряда Тейлора (3.15) с учетом возможно большего числа членов. При этом вторые и следующие производные рассчитываются не в результате дифференцирования, а путем многократного вычисления функции $f(x, u)$ в некоторых промежуточных точках между x_n и x_{n+1} . Проиллюстрируем основные идеи метода на примере получения схем второго порядка точности.

Оставим в разложении (3.15) члены вплоть до $O(h^2)$, имея в виду, что последний член разложения соответствует предполагаемому порядку точности схемы. Имеем

$$u_{n+1} = u_n + h u'_n + \frac{1}{2} h^2 u''_n. \quad (3.23)$$

Чтобы избежать явного дифференцирования, заменим u''_n в (3.23) разностью

$$u''_n = \frac{d}{dx} [f(x, u)] = \frac{f(\bar{x}, \bar{u}) - f(x_n, u_n)}{\Delta x}. \quad (3.24)$$

Величины $\bar{x}$, $\bar{u}$ и Δx подбираются так, чтобы обеспечить нужный порядок точности. Подставляя (3.24) в (3.23), получаем

$$u_{n+1} = u_n + h u'_n + \frac{1}{2} \frac{h^2}{\Delta x} [f(\bar{x}, \bar{u}) - f(x_n, u_n)]. \quad (3.25)$$

Полагая $\bar{x} = x_n + \gamma h$, $\bar{u} = u_n + \delta h$, обозначая $\alpha = (1/2)(h/\Delta x)$ и имея в виду, что $\Delta x = \bar{x} - x_n$, и, следовательно, $\gamma = 1/2\alpha$, а также учитывая, что $u'_n = f(x_n, u_n)$, переписываем (3.25) в виде

$$u_{n+1} = u_n + h \left[(1 - \alpha) f(x_n, u_n) + \alpha f(x_n + \frac{h}{2\alpha}, u_n + \delta h) \right]. \quad (3.26)$$

Параметры α и δ необходимо определить из условия, чтобы (3.26) наилучшим образом соответствовало ряду (3.15). Подразумевая, что

$$f(x_n + \frac{h}{2\alpha}, u_n + \delta h) = f(x_n, u_n) + \frac{h}{2\alpha} \frac{\partial f}{\partial x} + \delta h \frac{\partial f}{\partial u} \quad (3.27)$$

и подставляя (3.27) в (3.26), имеем

$$u_{n+1} = u_n + h f(x_n, u_n) + \frac{1}{2} h^2 \left(\frac{\partial f}{\partial x} + 2\delta \alpha \frac{\partial f}{\partial u} \right)_n. \quad (3.28)$$

Перепишем (3.15) в виде

$$u_{n+1} = u_n + h f(x_n, u_n) + \frac{1}{2} h^2 \left(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial u} \right)_n + O(h^3). \quad (3.29)$$

Из сравнения (3.27) и (3.29) очевидно, что первые три члена в этих формулах совпадают, если $\alpha\delta = (1/2)f(x_n, u_n)$. Выразая δ через α и подставляя в (3.26), получаем однопараметрическое семейство разностных схем Рунге-Кутты второго порядка точности. В результате

$$u_{n+1} = u_n + h \left\{ (1 - \alpha) f(x_n, u_n) + \alpha f \left(x_n + \frac{h}{2\alpha}, u_n + \frac{h}{2\alpha} f(x_n, u_n) \right) \right\}. \quad (3.30)$$

Схема (3.30) при $\alpha \neq 0$ имеет третий порядок точности на шаге и второй на интервале. Из (3.30) можно получить все рассмотренные выше разностные схемы. Так, при $\alpha = 0$ имеем метод Эйлера (см (3.16)), при $\alpha = 1/2$ – метод Эйлера с пересчетом (см. 3.(19)), а при $\alpha = 1$ – так называемый метод хорд. При $\alpha = 1$

$$u_{n+1} = u_n + h f \left(x_n + \frac{h}{2}, u_n + \frac{h}{2} f(x_n, u_n) \right). \quad (3.31)$$

Аналогичным образом могут быть получены и схемы Рунге-Кутты более высокого порядка точности. В настоящее время наиболее распространены схемы четвертого порядка точности. Например, схема Рунге-Кутты четвертого порядка точности, которая используется в большинстве стандартных программ на ЭВМ, выглядит так:

$$u_{n+1} = u_n + \frac{h}{6} (K_1 + 2K_2 + 2K_3 + K_4),$$

$$K_1 = f(x_n, u_n), \quad K_2 = f(x_n + \frac{h}{2}, u_n + \frac{h}{2} K_1),$$

$$K_3 = f(x_n + \frac{h}{2}, u_n + \frac{h}{2} K_2), \quad K_4 = f(x_n + h, u_n + h K_3).$$

3.6. МНОГОШАГОВЫЙ МЕТОД АДАМСА

В предыдущих схемах решение в точке x_{n+1} вычисляется с использованием решения только в одной точке x_n . Логично предположить, что можно повысить точность метода, если использовать информацию о поведении решения в предыдущих точках $x_n, x_{n-1}, \dots$. Такие методы получили название многошаговых.

Общая схема построения многошаговых методов выглядит следующим образом. Пусть нам известно приближенное решение в нескольких узлах сетки $x_n, x_{n-1}, \dots, x_{n-p}$. Следовательно, в этих точках отрезка известно и значение $f(x, u)$ правой части дифференциального уравнения (3.14) при $i = n, n-1, \dots, n-p$, причем $f(x, u(x))$ будет уже функцией только одной переменной x : $f(x, u(x)) = F(x)$. Заменяем функцию $F(x)$ интерполяционным многочленом Лагранжа $L_p(x)$ и вычислим значение u_{n+1} , проинтегрировав (3.14) на отрезке x_n, x_{n+1} . Находим

$$u_{n+1} = u_n + \int_{x_n}^{x_{n+1}} L_p(x) dx. \quad (3.32)$$

Проводя интегрирование, находим разностную схему для решения дифференциального уравнения. Порядок схемы определяется величиной остаточного члена интерполяционного полинома.

В случае, если для построения интерполяционного многочлена используются четыре узла $x_n, x_{n-1}, x_{n-2}, x_{n-3}$, то получается формула Адамса, которая на сетке с постоянным шагом записывается следующим образом:

$$u_{n+1} = u_n + \frac{h}{24} (55F_n - 59F_{n-1} + 37F_{n-2} - 9F_{n-3}) \quad (3.33)$$

с локальной ошибкой дискретизации

$$\varepsilon_n \leq \frac{251}{750} h^5 u^{(5)}(x),$$

т.е. метод имеет четвертый порядок точности на интервале.

Чтобы начать счет по схеме Адамса, необходимо знать решение в четырех начальных точках x_0, x_1, x_2, x_3 .

По существу, интерполяционный многочлен $L_p(x)$ в формуле (3.32) используется вне области интерполяции, т.е. в данном случае это экстраполяционный многочлен. Однако, поскольку интервал (x_n, x_{n+1}) мал, ошибка за счет экстраполяции невелика. Недостающие значения функции вычисляются в точках x_1, x_2, x_3 , как правило, по методу Рунге-Кутты соответствующего порядка. Это является недостатком метода, так как увеличивает объем программы на ЭВМ. Преимущество многошаговых методов заключается в том, что на каждом шаге правая часть дифференциального уравнения вычисляется только один раз, а в методе Рунге-Кутты четвертого порядка точности на каждом шаге функция $f(x, u)$ вычисляется четыре раза. Здесь, естественно, необходимо учитывать соотношение шагов двух методов, обеспечивающих заданную точность. В том случае, когда правая часть дифференциального уравнения является только функцией x , остаточный член в формуле Рунге-Кутты четвертого порядка точности на равномерной сетке равен остаточному члену формулы трапеций численного интегрирования с шагом $h/2$:

$$\varepsilon_n \leq \frac{b-a}{2880} h^4 M_4,$$

а в формуле Адамса

$$\varepsilon_n \leq \frac{b-a}{3} h^4 M_4.$$

Таким образом, остаточный член формулы Рунге-Кутты примерно на три порядка меньше, чем в формуле Адамса. Следовательно, для получения результата с одной и той же точностью можно шаг в методе Рунге-Кутты брать примерно в 6 раз больше, чем в формуле Адамса. Поэтому в методе Рунге-Кутты функция $f(x, u(x))$ вычисляется даже меньшее число раз.

Достоинство метода Адамса по сравнению с методом Рунге-Кутты заключается в простоте оценки остаточного члена метода.

3.7. НЕЯВНЫЕ СХЕМЫ. ПОНЯТИЕ О ЖЕСТКИХ СИСТЕМАХ

Рассмотрим следующую разностную схему для численного решения уравнения (3.14) при условии, что вычислено значение функции u в точке x_n . Имеем

$$u_{n+1} = u_n + h[sf(x_n, u_n) + (1-s)f(x_{n+1}, u_{n+1})], \quad (3.34)$$

где s – параметр разностной схемы, такой, что $0 \leq s \leq 1$. Говорят, что разностная схема является явной, если правая часть дифференциального уравнения записывается в известной точке (x_n, u_n) . Схема называется неявной, если правая часть дифференциального уравнения записывается в неизвестной точке (x_{n+1}, u_{n+1}) . Схема называется полуявной (полуявной), если правая часть с определенными весами записывается в известной и неизвестной точках. В соответствии с вышесказанным схема (3.34) является явной при $s = 1$, неявной при $s = 0$ и полуявной (полуявной) при $0 < s < 1$. В явных схемах решение в точке x_{n+1}, u_{n+1} определяется непосредственно из алгебраических соотношений с известными коэффициентами. Все рассмотренные выше схемы методов Эйлера, Рунге-Кутты, Адамса (кроме схемы метода Эйлера с пересчетом) являются явными. В неявных и полуявных схемах при $s \neq 1$ для нахождения решения в точке x_{n+1}, u_{n+1} необходимо решить в общем случае нелинейное уравнение (3.34).

Нелинейное уравнение (3.34) можно разрешить различными методами (метод простой итерации, метод дихотомии, метод Ньютона и т.д.) В рассматриваемом случае логично использовать метод простой итерации, поскольку нелинейное уравнение (3.34) уже разрешено относительно u_{n+1} . Тогда итерационный процесс можно записать в виде

$$u_{n+1}^{(k+1)} = u_n + h[sf(x_n, u_n) + (1-s)f(x_{n+1}, u_{n+1}^{(k)})], \quad (3.35)$$

где верхний индекс – номер итерации.

Известно, что сходимость метода простой итерации зависит от выбора начального приближения. В рассматриваемом случае в качестве $u_{n+1}^{(0)}$ можно выбрать u_n или какое-либо другое значение, например значение, вычисленное по методу Эйлера, которое достаточно близко к искомому значению u_{n+1} . Вычисления по формуле (3.35) продолжают до получения заданной точности $\varepsilon = |u^{(k+1)} - u^{(k)}|$.

Таким образом, очевидно, что при расчете на шаге явные схемы требуют значительно меньшего числа операций, чем неявные. Однако неявные схемы значительно более устойчивы, чем явные, в связи с чем в неявных схемах ограничения на шаг разностной схемы, связанные с устойчивостью, как правило, менее жесткие, чем в явных схемах.

В результате может оказаться, что общее число операций при заданной точности и при расчетах на интервале интегрирования в неявных схемах меньше, чем в явных.

В конечном итоге о качестве разностной схемы следует судить по числу операций, необходимых для получения решения с заданной точностью. Этот критерий и является основным при сравнении явных и неявных схем.

Понятие о жестких системах уравнений. Во многих задачах управления, химической кинетики, неравновесной динамики газов возникают системы урав-

нений, в которых содержатся некоторые уравнения, имеющие малый параметр при старшей производной. Соответственно матрица системы дифференциальных уравнений

$$\frac{du}{dx} = Au \quad (3.36)$$

имеет существенно различные собственные числа. Последнее обстоятельство приводит к тому, что в решении содержатся быстро и медленно убывающие члены, которые нужно рассчитывать с одинаковой точностью, что накладывает существенные ограничения на шаги разностной схемы. Дается следующее определение для жестких систем. Система называется жесткой, если действительная часть $\text{Re } \lambda_k$ меньше нуля:

$$\text{Re } \lambda_k < 0 \text{ и } l = \frac{\max |\text{Re } \lambda_k|}{\min |\text{Re } \lambda_k|} \gg 1,$$

где λ_k – собственные числа матрицы A , а l – так называемое число жесткости. Однако это определение не пригодно для одного дифференциального уравнения, так как для него $l = 1$. Поэтому в данном случае говорят, что уравнение является жестким, если содержит малый параметр при старшей производной. Ниже на примере модельного уравнения для химической кинетики будет показано, что численное решение жесткого уравнения рационально строить с использованием неявных или полунявных разностных схем, имея в виду условия устойчивости и ограничения на шаги.

Рассмотрим модельное жесткое линейное уравнение, характерное для задачи химической кинетики:

$$\frac{da}{dt} = \frac{1}{\tau}(\alpha^* - \alpha), \quad (3.37)$$

где $\alpha(t)$ – концентрация химического компонента; τ – время релаксации; α^* – равновесное значение химического компонента, достигаемое при $\tau \rightarrow 0$. Примем здесь, что τ и α^* – константы и τ мало, что характерно для условий приближения системы к равновесию.

Точное решение уравнения (3.37) есть

$$\alpha = (\alpha_0 - \alpha^*)e^{-\frac{t}{\tau}} + \alpha^*, \quad (3.38)$$

где $\alpha_0 = \alpha(0)$. Заменим (3.37) следующим разностным уравнением:

$$\frac{\alpha_{n+1} - \alpha_n}{h} = \frac{s}{\tau}(\alpha_0^* - \alpha_n) + \frac{1-s}{\tau}(\alpha^* - \alpha_{n+1}). \quad (3.39)$$

Из (3.39) следует рекуррентное соотношение

$$\alpha_{n+1} = \frac{1-\kappa s}{1+\kappa(1-s)}\alpha_n + \frac{\alpha^*\kappa}{1+\kappa(1-s)}, \quad \kappa = \frac{h}{\tau}. \quad (3.40)$$

Используя (3.40), получаем следующее точное решение разностного уравнения:

$$\alpha_{n+1} = (\alpha_0 - \alpha^*) \left[\frac{1-\kappa s}{1+\kappa(1-s)} \right]^{n+1} + \alpha^*, \quad (3.41)$$

а точное решение дифференциального уравнения в точке $n+1$ имеет вид

$$\alpha_{n+1} = (\alpha_0 - \alpha^*)e^{-\kappa(n+1)} + \alpha^*. \quad (3.42)$$

Из (3.41) и (3.42) следует, что решение разностного уравнения (3.41) стремится к точному при $\kappa \rightarrow 0$ для всех s . Однако порядок точности схемы и устойчивость решения разностного уравнения зависят от величины этого параметра. При $s = 1/2$ разностная схема является схемой второго порядка точности, а при остальных s – первого. При $s = 1$ (явная схема типа Эйлера или Рунге-Кутты) и $\kappa \gg 1$ (что имеет место, когда τ мало) решение разностного уравнения сильно отличается от точного.

Действительно, при $s = 1$ и $\kappa \gg 1$ член в квадратных скобках в уравнении (3.41) равен $(-\kappa)^{n+1}$, т.е. очень велик при больших κ , в то время как в точном решении соответствующий член $e^{-\kappa(n+1)}$ очень мал. Нетрудно видеть из (3.41), что максимальный шаг интегрирования при $s = 1$ для обеспечения устойчивого счета равен 2τ . Для задач химической кинетики характерны $\tau = 10^{-6} - 10^{-9}$. Поэтому явные схемы позволяют численно решать жесткие уравнения лишь с очень малым шагом $h \approx 2\tau$, что делает их абсолютно непригодными даже при использовании ЭВМ с большим быстродействием.

В связи с этим более предпочтительной является неявная схема с $s = 0$, которая при $\kappa \gg 1$ дает решения, близкие к точному. Действительно, при $s = 0$ член в квадратных скобках в уравнении (3.41) равен $[1/\kappa]^{n+1}$ и качественно верно отражает поведение экспоненты $e^{-\kappa(n+1)}$ в точном решении. Применение неявной схемы позволяет существенно увеличить шаг интегрирования по сравнению с явными схемами при сохранении устойчивости и необходимой точности расчета.

3.8. ПОСТАНОВКА КРАЕВОЙ ЗАДАЧИ ДЛЯ ОБЫКНОВЕННОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ

Краевая задача – это задача отыскания частного решения системы обыкновенных дифференциальных уравнений

$$\frac{du_i}{dx} = f_i(x, u_1, u_2, \dots, u_n), \quad i = 1, 2, \dots, n \quad (3.43)$$

с дополнительными условиями, налагаемыми на значения функций $u_i(x)$ не менее чем в двух точках отрезка $[a, b]$. Следовательно, краевая задача для ОДУ

ставится для системы дифференциальных уравнений порядка не менее второго (или одного дифференциального уравнения порядка не ниже второго).

Свое название краевая задача исторически получила по тому случаю, когда дополнительные условия заданы на концах (краях) отрезка $[a, b]$. Естественно, дополнительные условия могут задаваться и во внутренних точках отрезка. Такие условия называются внутренними краевыми условиями. Краевые условия могут связывать между собой значения нескольких функций, производных функций или комбинаций функций и производных в одной или нескольких точках отрезка, на котором ищется решение.

Точное аналитическое решение краевой задачи удастся найти крайне редко, так как для этого надо найти общее решение системы дифференциальных уравнений (3.43) и выразить из краевых условий значения входящих в него констант. Поэтому широкое распространение получили численные методы решения краевых задач, такие как метод стрельбы и разностный метод.

3.9. МЕТОД СТРЕЛЬБЫ

Рассмотрим метод стрельбы на примере решения дифференциального уравнения второго порядка

$$u'' = f(x, u, u'), \quad x \in [a, b] \quad (3.44)$$

с краевыми условиями:

$$u(a) = \alpha, \quad (3.44a)$$

$$u(b) = \beta. \quad (3.44b)$$

Основная идея метода стрельбы заключается в сведении решения краевой задачи (3.44), (3.44a), (3.44b) для того же дифференциального уравнения (3.44) к решению серии задач Коши. Для постановки задачи Коши для уравнения (3.44) необходимо в какой-либо одной точке отрезка задать два дополнительных условия. В точке a известно одно дополнительное условие $u(a) = \alpha$. Поэтому зададим в этой точке значение производной функции – $u'(a)$. Так как это значение заранее неизвестно, то зададим его равным некоторому произвольному значению η . В результате получим задачу Коши:

$$u'' = f(x, u, u'), \quad (3.45)$$

$$u(a) = \alpha, \quad (3.45a)$$

$$u'(a) = \eta. \quad (3.45b)$$

Решая эту задачу Коши каким-либо численным методом, получаем ее решение $u(x, \eta)$, зависящее от η , как от параметра. Так как η выбрано произвольно, то решение задачи Коши удовлетворяет условию краевой задачи в точке a и не удовлетворяет ее условию в точке b . Таким образом, необходимо менять параметр η таким образом, чтобы решение задачи Коши в точке b совпадало с

условием (3.44b). Следовательно, решение краевой задачи сводится к нахождению корня нелинейного алгебраического уравнения

$$(\eta) = u(b, \eta) - \beta = 0. \quad (3.46)$$

При этом функция (η) задана не аналитически, а в виде таблицы чисел, которая, очевидно, составляется при решении серии задач Коши. Решение уравнения (3.46) можно искать методом дихотомии. Делают пробные «выстрелы», т.е. решают задачу Коши с разными значениями условия η , до тех пор, пока среди величин $u(b, \eta) - \beta$ не окажется двух разных по знаку. Пара соответствующих значений η делится пополам, и находится новое значение η , с которым решается задача Коши. Такая процедура повторяется до получения условия (3.45b) с необходимой точностью.

Для ускорения сходимости к корню уравнения (3.46) можно использовать другие методы, например метод секущих. Для этого делают два расчета с произвольными значениями $\eta^{(0)}$ и $\eta^{(1)}$, а следующие значения η вычисляют по формуле

$$\eta^{(k+1)} = \eta^{(k)} - \frac{(\eta^{(k)} - \eta^{(k-1)}) [u(b, \eta^{(k)}) - \beta]}{u(b, \eta^{(k)}) - u(b, \eta^{(k-1)})}.$$

Простота алгоритма метода стрельбы и возможность использования стандартных программ решения задачи Коши позволяют успешно использовать его при решении как линейных, так и нелинейных дифференциальных уравнений.

3.10. КОНЕЧНО-РАЗНОСТНЫЙ МЕТОД РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ

Идея метода заключается в сведении краевой задачи к решению системы алгебраических уравнений путем замены производных в дифференциальном уравнении и краевых условий конечно-разностными соотношениями. Рассмотрим идею метода на примере краевой задачи для уравнения второго порядка:

$$u'' + A(x)u' + B(x)u = f(x), \quad (3.47)$$

$$u(a) = \alpha, \quad (3.47a)$$

$$u(b) = \beta. \quad (3.47b)$$

Введем на $[a, b]$ сетку $a = x_0 < x_1 < \dots < x_n = b$. Для простоты будем полагать шаг сетки постоянным. Будем искать решение не на всем отрезке $[a, b]$, а лишь в $(N-1)$ узлах сетки. Таким образом, нам необходимо найти $(N-1)$ неизвестных значений функции u в узлах сетки. Воспользуемся для аппроксимации производных функций формулами численного дифференцирования (см. гл. 2):

$$u' \approx \frac{u_{i+1} - u_{i-1}}{2h} + O(h^2), \quad u'' \approx \frac{1}{h^2}(u_{i-1} - 2u_i + u_{i+1}) + O(h^2), \quad i = 1, 2, \dots, N-1. \quad (3.48)$$

Такую аппроксимацию можно записать для каждого внутреннего узла сетки и получить систему линейных уравнений $(N-1)$ порядка, если подставить (3.48) в (3.47). Два недостающих уравнения получаем из краевых условий (3.47а) и (3.47б).

Запишем эту линейную систему в виде

$$\begin{aligned} a_i u_{i-1} - b_i u_i + c_i u_{i+1} &= d_i, \quad i = 1, 2, \dots, N-1, \\ u_0 &= \alpha, \\ u_n &= \beta, \end{aligned} \quad (3.49)$$

где $a_i = \left(\frac{1}{h^2} - \frac{A_i}{2h} \right)$, $b_i = \left(\frac{2}{h^2} - B_i \right)$, $c_i = \left(\frac{1}{h^2} + \frac{A_i}{2h} \right)$.

Нетрудно видеть, что матрица этой линейной системы – трехдиагональная и поэтому решать ее можно методом прогонки (см. гл. 1). При этом, как обычно, на прямом ходе определяются прогоночные коэффициенты, а на обратном – неизвестные u_i . Устойчивость и корректность прогонки, как известно, обеспечивается условием преобладания диагональных элементов.

Описанный метод можно применить и в случае, если в краевые условия (3.47а) и (3.47б) входят линейные комбинации функций и первых производных.

Глава 4. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

4.1. КРАТКОЕ СОДЕРЖАНИЕ ГЛАВЫ

Уравнения в частных производных описывают многие физические задачи в таких областях, как механика сплошных сред, термодинамика, квантовая механика, электродинамика, теория упругости и многие другие. Аналитическое решение уравнения в частных производных удается получить лишь в единичных практически важных случаях, и поэтому роль численных методов для решения задач, описываемых уравнениями в частных производных, особенно велика.

В этой главе излагаются основные методы численного решения уравнений в частных производных, такие как метод сеток, метод прямых и метод характеристик, метод установления, методы сквозного счета, а также некоторые специальные методы, нашедшие применение в последнее время.

Особое внимание уделяется методу сеток как наиболее развитому и употребительному методу решения уравнений в частных производных. При изложении метода сеток вводятся такие основные понятия, как аппроксимация, устойчивость и сходимость разностной схемы, имеющие первостепенное значение при рассмотрении любого численного метода.

Представлены также некоторые сведения из теории уравнений в частных производных.

4.2. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Постановку задач для уравнений математической физики дадим на основе классических уравнений параболического, гиперболического и эллиптического типов

1. Основные определения. Если искомая функция зависит не от одной, а от нескольких переменных, то при рассмотрении многих важных задач возникают уравнения с частными производными, связывающие между собой независимые переменные $x_1, x_2, \dots, x_k$, а также искомую функцию $u(x_1, x_2, \dots, x_k)$ и её частные производные $\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_k}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \frac{\partial^2 u}{\partial x_2^2}, \frac{\partial^2 u}{\partial x_2 \partial x_3}$ и т.д.

В общем случае уравнение в частных производных может быть записано в следующем виде

$$F\left(x_1, \dots, x_k, u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_k}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \dots, \frac{\partial^n u}{\partial x_1^{p_1} \partial x_2^{p_2} \dots \partial x_k^{p_k}}\right) = 0,$$

где $p_1 + p_2 + \dots + p_k = n$.

Наивысший порядок входящей в дифференциальное уравнение производной называется порядком дифференциального уравнения.

Если функция F линейна относительно функции u и её производных, то приведенное выше уравнение называется линейным уравнением в частных производных. Если $F = L(u) = f$ – линейное дифференциальное уравнение, то уравнение $L(u) = f(x_k)$ называется неоднородным, а $L(u) = 0$ – линейным однородным дифференциальным уравнением.

Если функция F линейна по высшим производным (n -го порядка), т.е. коэффициенты при высших производных могут зависеть лишь от функции u и производных до $(n-1)$ -го порядка, то дифференциальное уравнение называется квазилинейным.

Классическим решением называется функция $u(x_1, \dots, x_k)$, имеющая частные производные до требуемого порядка и обращающая приведенное выше уравнение в тождество. Решение уравнений с частными производными, как правило, не единственно. Если в случае ОДУ общее решение зависит от конечного числа параметров, то для уравнений в частных производных обычно имеется общее решение, зависящее от производных функций с числом аргументов на единицу меньше, чем у функции $u(x_1, \dots, x_k)$.

Например, уравнение

$$\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2} = 0$$

имеет решение, зависящее от произвольной функции

$$u(x_1, x_2) = \varphi(x_1 - x_2).$$

Таким образом, так же, как и для ОДУ, для выделения единственного решения необходимо задавать дополнительные условия, куда будут входить функции с числом аргументов $k-1$. В качестве таких условий можно потребовать, чтобы искомая функция $u(x_1, \dots, x_k)$ принимала заданные значения на поверхности размерности $k-1$ в пространстве переменных $x_1, x_2, \dots, x_k$. К этому типу условий относятся начальные и граничные условия, являющиеся наиболее часто встречающимися при постановке краевых задач математической физики.

2. Уравнения математической физики. Ниже будут рассмотрены в основном классические уравнения математической физики. В этих уравнениях в качестве независимых переменных используются время t и пространственные координаты: декартовы, цилиндрические и сферические.

Задача называется стационарной, если решение не зависит от времени, и нестационарной или эволюционной, если оно зависит от времени. Задачи с од-

ной пространственной переменной называются одномерными, а с двумя – двумерными, с тремя – трехмерными.

Запишем каноническую форму уравнений с двумя независимыми переменными. Заметим, что в общем случае для более чем двух независимых переменных таких канонических форм не существует.

Имеем

$$A(x, y) \frac{\partial^2 u}{\partial x^2} + 2B(x, y) \frac{\partial^2 u}{\partial x \partial y} + C(x, y) \frac{\partial^2 u}{\partial y^2} + F\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) = 0$$

с дважды непрерывно дифференцируемыми коэффициентами A, B, C , которые не обращаются в нуль одновременно. Имеем следующую классификацию дифференциальных уравнений в частных производных. Пусть $D = B^2 - AC$, тогда при $D \equiv 0$ имеем параболическое, при $D > 0$ – гиперболическое, а при $D < 0$ – эллиптическое дифференциальное уравнение. При этом тип одного и того же уравнения может меняться в зависимости от области пространства.

Классическим уравнением гиперболического типа является волновое уравнение параболического типа – уравнение теплопроводности, а эллиптического типа – уравнение Лапласа.

В математической физике различают следующие задачи: задачу Коши (или задача с начальными данными), краевые или граничные задачи, смешанные краевые задачи или краевые (граничные) задачи с начальными данными. Задачи, в которых в качестве переменной имеется время t , иногда называют эволюционными.

В теории уравнений в частных производных важную роль играют характеристические поверхности, которые определяются как поверхности, в окрестности которых решения задачи Коши не существует, либо решение это не единственно. Можно показать, что в действительной плоскости характеристические поверхности существуют для гиперболических и параболических уравнений и не существуют для эллиптических уравнений.

Общая теория характеристик позволяет сформулировать для гиперболических уравнений некоторые специфические задачи, такие, как задача Гурса, смешанная задача и ряд других. Их решение можно построить, используя как аналитический, так и численный метод характеристик.

3. Постановка задач для уравнений параболического типа. Классическим примером уравнения параболического типа является уравнение теплопроводности или диффузии в области $[0, l] + [0, t]$ (рис. 4.1):

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < l,$$

где a – коэффициент теплопроводности (если u – температура) и массопроводности (если u – концентрация, давление в задачах фильтрации и т.п.). Поскольку имеется производная по вре-

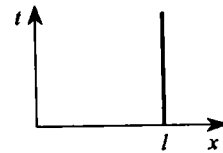


Рис. 4.1

мени, то необходимо задавать начальные условия при $t=0$ и граничные условия при $x=0, x=l, t>0$

Для уравнения теплопроводности при краевых (граничных) условиях

$$u(0, t) = \varphi_1(t), \quad x = 0, \quad t > 0,$$

$$u(l, t) = \varphi_2(t), \quad x = l, \quad t > 0$$

и начальном условии

$$u(x, 0) = \psi(x), \quad 0 \leq x \leq l, \quad t = 0$$

имеем *первую* начально-краевую или краевую задачу с начальными данными – (смешанную) задачу, иначе, задачу Коши с краевыми (граничными) условиями

Для уравнения теплопроводности с краевыми условиями

$$\frac{\partial u(0, t)}{\partial x} = \varphi_1(t), \quad x = 0, \quad t > 0,$$

$$\frac{\partial u(l, t)}{\partial x} = \varphi_2(t), \quad x = l, \quad t > 0$$

и начальными данными

$$u(x, 0) = \psi(x), \quad 0 \leq x \leq l, \quad t = 0$$

имеем *вторую* краевую задачу с начальными данными (смешанную задачу)

Для уравнения теплопроводности с краевыми условиями

$$\alpha_0 \frac{\partial u(0, t)}{\partial x} + \beta_0 u(0, t) = \varphi_1(t), \quad x = 0, \quad t > 0,$$

$$\alpha_1 \frac{\partial u(l, t)}{\partial x} + \beta_1 u(l, t) = \varphi_2(t), \quad x = l, \quad t > 0$$

и начальными условиями

$$u(x, 0) = \psi(x)$$

имеем *третью* краевую задачу с начальными данными (смешанную задачу)

Для бесконечной области ставится следующая *задача Коши* для уравнения теплопроводности

$$u(x, 0) = \varphi(x), \quad -\infty < x < +\infty, \quad t = 0$$

4 Постановка задач для уравнений гиперболического типа. Классическим примером уравнения гиперболического типа является *волновое уравнение*

$$\frac{\partial^2 u}{\partial t^2} = a \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < l, \quad t > 0,$$

или более общее, описывающее малые продольные колебания стержня и поперечные колебания струны, причем функция $u(x, t)$ – отклонение струны от по-

ложения равновесия и a – скорость перемещения колебаний. К волновому уравнению сводятся также уравнения распространения малых акустических возмущений

Поскольку в волновое уравнение входит вторая производная по времени, то старший порядок производной по времени в начальных условиях не превышает единицы. Тогда первая, вторая и третья начально-краевые или краевые задачи с начальными данными для волнового уравнения будут иметь соответственно следующий вид

$$\begin{cases} u(0, t) = \varphi_1(t), \quad x = 0, \quad t > 0, \\ u(l, t) = \varphi_2(t), \quad x = l, \quad t > 0, \\ u(x, 0) = \psi_1(x), \quad 0 \leq x \leq l, \quad t = 0, \\ \frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad 0 \leq x \leq l, \quad t = 0, \end{cases}$$

$$\begin{cases} \frac{\partial u(0, t)}{\partial x} = \varphi_1(t), \quad x = 0, \quad t > 0, \\ \frac{\partial u(l, t)}{\partial x} = \varphi_2(t), \quad x = l, \quad t > 0, \\ u(l, t) = \psi_1(t), \quad x = l, \quad t > 0, \\ u(x, 0) = \psi_2(x), \quad 0 \leq x \leq l, \quad t = 0, \end{cases}$$

$$\begin{cases} \alpha_0 \frac{\partial u(0, t)}{\partial x} + \beta_0 u(0, t) = \varphi_1(t), \quad x = 0, \quad t > 0, \\ \alpha_1 \frac{\partial u(l, t)}{\partial x} + \beta_1 u(l, t) = \varphi_2(t), \quad x = l, \quad t > 0, \\ u(x, 0) = \psi_1(x), \quad 0 \leq x \leq l, \quad t = 0, \\ \frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad 0 \leq x \leq l, \quad t = 0 \end{cases}$$

В бесконечной области можно поставить задачу Коши

$$\begin{cases} u(x, 0) = \psi_1(x), \quad -\infty < x < +\infty, \quad t = 0, \\ \frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad -\infty < x < +\infty, \quad t = 0 \end{cases}$$

Отметим, что метод характеристик позволяет решать задачу Коши, Гурса и смешанные задачи при задании данных в ограниченной области

5 Постановка задач для уравнений эллиптического вида. Классическим примером уравнения эллиптического типа является *уравнение Лапласа*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad x, y \in \Omega,$$

или уравнение Пуассона

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y),$$

описывающее течение идеальной (без вязкости и теплопроводности) жидкости в стационарных потоках, стационарное распределение температуры, стационарное распределение напряженности электрического или магнитного полей. При этом уравнение Лапласа описывает все эти явления, когда нет источников или стоков (нет правых частей), а уравнение Пуассона – с распределенными по области Ω источниками, задаваемыми правой частью $f(x, y)$ (рис. 4.2).

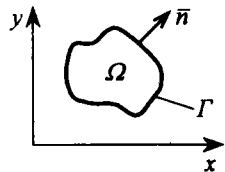


Рис. 4.2

Поскольку уравнение Лапласа (Пуассона) – стационарное, то начальные условия не задаются; на границах же Γ расчетной области $\Omega + \Gamma$ задаются граничные условия 1, 2 и 3-го родов.

Первая краевая задача для уравнения Лапласа (задача Дирихле):

$$u|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma.$$

Вторая краевая задача для уравнения Лапласа (задача Неймана):

$$\frac{\partial u}{\partial n}|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma.$$

Третья краевая задача для уравнения Лапласа:

$$\alpha \frac{\partial u}{\partial n}|_{\Gamma} + \beta u|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma.$$

При этом во второй и третьей краевой задаче производные берутся в направлении внешней нормали по отношению к области Ω .

6. Постановка многомерных задач математической физики. В заключение приведем примеры постановок многомерных по пространственным переменным задач для уравнений параболического и эллиптического типов (рис. 4.3)

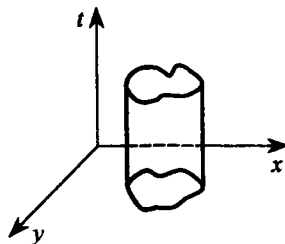


Рис. 4.3

Первая начально-краевая задача для двумерного уравнения теплопроводности:

$$\begin{cases} \frac{\partial u}{\partial t} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, \quad t > 0, \\ u|_{\Gamma} = \varphi(x, y, t), & (x, y) \in \Gamma, \quad t > 0, \\ u(x, y, 0) = \psi(x, y), & (x, y) \in \Omega + \Gamma, \quad t = 0. \end{cases}$$

Вторая начально-краевая задача для двумерного уравнения теплопроводности:

$$\begin{cases} \frac{\partial u}{\partial t} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, \quad t > 0, \\ \frac{\partial u}{\partial n}|_{\Gamma} = \varphi(x, y), & (x, y) \in \Gamma, \quad t > 0, \\ u(x, y, 0) = \psi(x, y, t), & (x, y) \in \Omega + \Gamma, \quad t = 0. \end{cases}$$

Третья начально-краевая задача для двумерного уравнения теплопроводности:

$$\begin{cases} \frac{\partial u}{\partial t} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, \quad t > 0, \\ \alpha \frac{\partial u}{\partial n}|_{\Gamma} + \beta u|_{\Gamma} = \varphi(x, y, t), & (x, y) \in \Gamma, \quad t > 0, \\ u(x, y, 0) = \psi(x, y), & (x, y) \in \Omega + \Gamma, \quad t = 0. \end{cases}$$

Первая начально-краевая задача для двумерного волнового уравнения:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, \quad t > 0, \\ u|_{\Gamma} = \varphi(x, y, t), & (x, y) \in \Gamma, \quad t > 0, \\ u(x, y, 0) = \psi_1(x, y), & (x, y) \in \Omega + \Gamma, \quad t = 0, \\ \frac{\partial u(x, y, 0)}{\partial t} = \psi_2(x, y), & (x, y) \in \Omega + \Gamma, \quad t = 0. \end{cases}$$

Вторая начально-краевая задача для двумерного волнового уравнения:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, t > 0, \\ \left. \frac{\partial u}{\partial n} \right|_{\Gamma} = \varphi(x, y), & (x, y) \in \Gamma, t > 0, \\ u(x, y, 0) = \psi_1(x, y), & (x, y) \in \Omega + \Gamma, t = 0, \\ \left. \frac{\partial u(x, y, 0)}{\partial t} \right| = \psi_2(x, y), & (x, y) \in \Omega + \Gamma, t = 0. \end{cases}$$

Третья начально-краевая задача для двумерного волнового уравнения:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = a \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), & (x, y) \in \Omega, t > 0, \\ \alpha \left. \frac{\partial u}{\partial n} \right|_{\Gamma} + \beta u|_{\Gamma} = \varphi(x, y, t), & (x, y) \in \Gamma, t > 0, \\ u(x, y, 0) = \psi_1(x, y), & (x, y) \in \Omega + \Gamma, t = 0, \\ \left. \frac{\partial u(x, y, 0)}{\partial t} \right| = \psi_2(x, y), & (x, y) \in \Omega + \Gamma, t = 0. \end{cases}$$

4.3. ОСНОВНЫЕ ПОНЯТИЯ МЕТОДА СЕТОК. ЗАДАЧА ДИРИХЛЕ ДЛЯ УРАВНЕНИЯ ЛАПЛАСА

Метод сеток (или метод конечных разностей) сводит решение систем уравнений в частных производных к решению систем, как правило, линейных алгебраических уравнений с достаточно разреженными матрицами. При этом построение решения в методе сеток осуществляется в три этапа.

1. Область непрерывного изменения аргумента (или аргументов) заменяется конечным дискретным множеством точек, называемых разностной сеткой. В разностной сетке выделяются внутренние и граничные узлы. Решение ищется во внутренних узлах, а в граничных узлах значение искомой функции задается при аппроксимации граничных условий исходной дифференциальной задачи. Функция дискретного аргумента, определенная на разностной сетке, называется сеточной функцией.

2. Дифференциальное уравнение и граничные условия заменяются по определенным правилам своими разностными аналогами. Разностные операторы, соответствующие дифференциальному уравнению, записываются во внутренних узлах сетки. Разностные операторы, соответствующие граничным условиям, записываются в граничных узлах. В результате получается система алгебраических уравнений, число которых пропорционально числу внутренних узлов разностной сетки.

3. Осуществляется решение системы алгебраических уравнений каким-либо из известных методов. В большинстве случаев получаемая система урав-

нений есть система линейных алгебраических уравнений очень большого порядка, но с очень разреженной матрицей. В случае нелинейных систем итерационные процедуры, как правило, сводят их к линейным системам.

Основные понятия и этапы метода сеток продемонстрируем на примере решения задачи Дирихле для классического уравнения эллиптического типа – уравнения Лапласа. Итак, имеем уравнение Лапласа

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad (4.1)$$

и граничное условие

$$u(x, y)|_{\Gamma} = u_0(x, y), \quad (4.2)$$

где Γ – граница области G (рис. 4.4), в которой ищется решение $u(x, y)$, удовлетворяющее уравнению (4.1) и граничному условию (4.2).

На первом этапе метода сеток область G непрерывного изменения аргумента с границей Γ заменяют приближающей её сеточной областью G_h с границей Γ_h . Для этого проведем линии $x = \text{const}$ и $y = \text{const}$, так что $x = mh_x$ ($m = \overline{0, M}$) и $y = nh_y$ ($n = \overline{0, N}$). Величины h_x и h_y , называемые шагами разностной сетки, в общем случае могут быть различными.

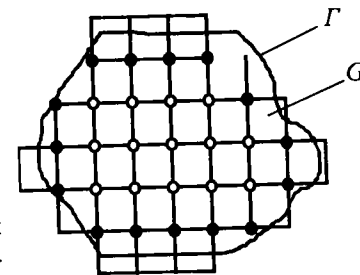


Рис. 4.4

Точки пересечения линий $x = \text{const}$ и $y = \text{const}$ называют узлами разностной сетки. Различают два типа узлов – внутренние и внешние. Внутренними называют такие узлы, для которых четыре соседних узла (по два в каждом направлении) принадлежат области $\bar{G} = G + \Gamma$. Сеточной функции приписываются нижние индексы. На рис. 4.4 внутренние узлы обозначены пустыми кружочками, а граничные – заштрихованными.

На втором этапе заменим дифференциальный оператор Лапласа разностным оператором. С этой целью выберем шаблон разностной схемы – набор (конфигурацию) узлов, с использованием которых производится замена производных конечными разностями. Шаблон, содержащий p точек, называется p -точечным. Для аппроксимации вторых производных, входящих в оператор Лапласа, используем пятиточечный шаблон, показанный на рис. 4.5. Не теряя общности рассуждений, рассмотрим далее случай, когда шаги разностной сетки по на-

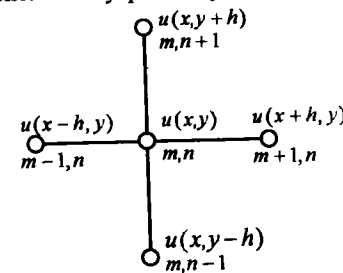


Рис. 4.5

правлению x и y одинаковы, т.е. $h_x = h_y = h$. Используя разложение функции на точном решении в ряд Тейлора в окрестности точки (x, y) , имеем:

$$u(x \pm h, y) = u(x, y) \pm h \frac{\partial u}{\partial x} + \frac{h^2}{2!} \frac{\partial^2 u}{\partial x^2} \pm \frac{h^3}{3!} \frac{\partial^3 u}{\partial x^3} + O(h^4),$$

$$u(x, y \pm h) = u(x, y) \pm h \frac{\partial u}{\partial y} + \frac{h^2}{2!} \frac{\partial^2 u}{\partial y^2} \pm \frac{h^3}{3!} \frac{\partial^3 u}{\partial y^3} + O(h^4).$$

Из этих соотношений нетрудно получить аппроксимации вторых производных со вторым порядком точности:

$$\frac{\partial^2 u}{\partial x^2} = \frac{u(x+h, y) - 2u(x, y) + u(x-h, y)}{h^2} + O(h^2),$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{u(x, y+h) - 2u(x, y) + u(x, y-h)}{h^2} + O(h^2).$$

Введем для сокращения записи общепринятые индексные обозначения узлов в соответствии с рис. 4.5 и соответствующие обозначения сеточных функций в узлах: $u_{m,n}$, $u_{m-1,n}$ и т.д. Тогда соотношения для производных запишем в виде

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{m+1,n} - 2u_{m,n} + u_{m-1,n}}{h^2} + O(h^2), \quad (4.3)$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{u_{m,n+1} - 2u_{m,n} + u_{m,n-1}}{h^2} + O(h^2). \quad (4.4)$$

Используя соотношение (4.3) и (4.4), записываем разностный оператор Лапласа на пятиточечном шаблоне следующим образом:

$$\Delta u_h = \frac{u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1} - 4u_{m,n}}{h^2} + O(h^2). \quad (4.5)$$

Для аппроксимации граничных условий используются различные приемы. Для случая прямоугольной области с границами, совпадающими с линиями $x = \text{const}$ и $y = \text{const}$, граничные значения известны в узлах точно, так как $u_{0h} = u_0(mh_x, nh_y)$. В случае области произвольной формы значения сеточной функции в граничных узлах принимаются равными значениям функции u_0 в ближайших (по какому-либо критерию) точках кривой Γ (см. рис. 4.4). В случае задачи Неймана или смешанной задачи, когда на границе области задается значение производной функции u , производится аппроксимация производной

Записав разностный оператор (4.5) во всех внутренних узлах, получим систему «сеточных» уравнений для нахождения значений сеточной функции $u_{m,n}$ в узлах разностной сетки.

На третьем этапе осуществляется решение системы сеточных уравнений. В общем случае сеточные уравнения могут быть нелинейными. Однако в рассматриваемой задаче, в силу линейности основного дифференциального уравнения Лапласа, система уравнений (4.5) является системой линейных алгебраических уравнений для нахождения значений неизвестных во внутренних узлах. При этом число уравнений системы точно равно числу внутренних узлов разностной сетки. Отметим, что число уравнений может быть весьма велико. Так, для нахождения достаточно точного решения рассматриваемой задачи требуется задать $N = M \approx 50 + 100$, и, следовательно, количество уравнений достигает нескольких сотен или даже тысяч.

Перепишем систему сеточных решений в виде

$$u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1} - 4u_{m,n} = 0. \quad (4.6)$$

Отметим, что каждое из уравнений системы содержит лишь пять неизвестных $u_{m+1,n}$, $u_{m-1,n}$, $u_{m,n+1}$, $u_{m,n-1}$, $u_{m,n}$, хотя в полной системе таких неизвестных $\approx N^2$. Таким образом, матрица системы (4.6) является сильно разреженной. Для решения систем с сильно разреженными матрицами хорошо приспособлены итерационные методы решения систем линейных уравнений – метод простой итерации, метод итерации Зейделя, методы релаксации. Рассмотрим их использование для решения системы (4.6).

Систему (4.6) запишем в форме

$$u_{m,n} = \frac{1}{4} (u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1}). \quad (4.7)$$

Видим, что значение функции в центральной точке есть среднее арифметическое значений функции в четырех соседних по направлениям x и y узлах. Уравнение (4.7) можно интерпретировать как разностный аналог теоремы о среднем для гармонических функций.

Рассмотрим применение метода простой итерации для решения системы (4.7). Пусть область G , в которой разыскивается решение, имеет для простоты прямоугольную форму (рис. 4.6). Построим итерационный процесс

$$u_{m,n}^{(k)} = \frac{1}{4} (u_{m+1,n}^{(k-1)} + u_{m-1,n}^{(k-1)} + u_{m,n+1}^{(k-1)} + u_{m,n-1}^{(k-1)}), \quad (4.8)$$

где верхний индекс означает номер итерации. В граничных точках значения известны точно. Для нахождения решения по соотношению (4.8) необходимо знать начальное приближение – значения функции на нулевой итерации $u_{m,n}^{(0)}$. Доказывается, что для любого шага разностной сетки итерационный процесс (4.8) сходится независимо от начального приближения $u_{m,n}^{(0)}$. На практике желательно «выбирать» начальное приближение возможно более близким к точному решению. Например, на основании теоремы о среднем для гармонических функций в качестве «хорошего» начального приближения можно принять зна-

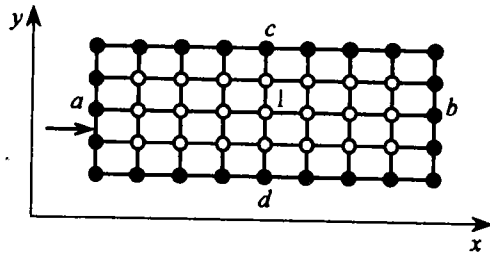


Рис. 4.6

чение функции, полученное интерполяцией на область G значений функции в граничных узлах. Так, значение функции $u^{(0)}$ в «центральной» точке l (рис. 4.6) может быть взято как среднее арифметическое значений функции в граничных узлах a, b, c, d .

Сходимость итерационного процесса можно улучшить, если использовать итерационный процесс Зейделя, например, в виде

$$u_{m,n}^{(k)} = \frac{1}{4} \left(u_{m+1,n}^{(k-1)} + u_{m-1,n}^{(k-1)} + u_{m,n+1}^{(k-1)} + u_{m,n-1}^{(k-1)} \right). \quad (4.9)$$

Члены с индексами $(m-1, n)$ и $(m, n-1)$ берутся с k -й итерации, т.е. в вычислениях используются уточненные значения функции в этих точках. Как легко понять из рис. 4.6, расчет по формуле (4.9) при движении, например, слева направо от границы области позволяет последовательно вычислить значения функции на слоях $x = \text{const}$ (или $y = \text{const}$). Использование уточненных значений функции улучшает сходимость метода итераций и позволяет уменьшить требуемый для реализации метода объем оперативной памяти ЭВМ, так как в методе Зейделя не требуется одновременно хранить значение функции в каждой точке на двух итерациях. Итерации заканчиваются при выполнении условия $\max_{m,n} |u_{m,n}^{(k)} - u_{m,n}^{(k-1)}| \leq \varepsilon$, где ε – заданная погрешность определения решения.

В настоящее время распространены релаксационные методы решения системы (4.7), в которых итерационный процесс строится по формуле $u_{m,n}^{(k)} = \omega u_{m,n}^{(k-1)} + \frac{1-\omega}{4} (u_{m+1,n}^{(k-1)} + u_{m-1,n}^{(k-1)} + u_{m,n+1}^{(k-1)} + u_{m,n-1}^{(k-1)})$, где ω – параметр релаксации. В методе релаксации член $\omega u_{m,n}^{(k-1)}$ учитывает информацию о величине $u_{m,n}$ из предыдущей итерации. При $\omega > 1$ итерационный метод называется методом верхней релаксации, при $\omega = 1$ – методом полной релаксации и при $\omega < 1$ – методом нижней релаксации.

Для рассматриваемой нами задачи можно показать, что число итераций I , необходимых для достижения заданной точности, определяется соотношениями, приведенными в табл. 4.1.

Таблица 4.1

Название метода	Число итераций I
Метод простой итерации	$0.2 N^2 \ln \varepsilon^{-1}$
Метод итераций Зейделя	$0.1 N^2 \ln \varepsilon^{-1}$
Метод верхней релаксации	$0.7 N \ln \varepsilon^{-1}$

Видно, что в методе верхней релаксации число итераций для достижения заданной точности линейно зависит от числа узлов N , тогда как в остальных методах зависимость квадратичная.

4.4. ЯВНЫЕ И НЕЯВНЫЕ РАЗНОСТНЫЕ СХЕМЫ

При решении задач для уравнений параболического и гиперболического типов используются различные разностные схемы, среди которых важное место занимают так называемые явные и неявные разностные схемы.

Рассмотрим их на примере классического уравнения параболического типа – уравнения теплопроводности (нестационарного уравнения Лапласа)

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad (4.10)$$

со следующими начальными

$$u(x, y, 0) = u_0(x, y) \quad (4.11)$$

и граничными

$$u(x, y, t)|_{\Gamma} = u^0(x, y, t) \quad (4.12)$$

условиями. Решение ищется внутри полубесконечного цилиндра (рис. 4.7). Задачу (4.10) – (4.12) называют эволюционной, имея в виду построение эволюции решения во времени. Используются также названия: задача Коши с граничными условиями, или краевая задача с начальными данными.

Построим разностную схему. Для этого проведем линии $x = m h_x$ ($m = \overline{0, M}$), $y = n h_y$ ($n = \overline{0, N}$) и $t = k \tau$ ($k = \overline{0, K}$), где h_x и h_y – шаги соответственно по x и y (по пространственным координатам) и τ – шаг по t (по времени). Плоскости $k = \text{const}$ называют t -слоями или слоями по времени. Сеточную функцию $u(x + m h_x, y + n h_y, t + k \tau)$ обозначим $u_{m,n}^k$.

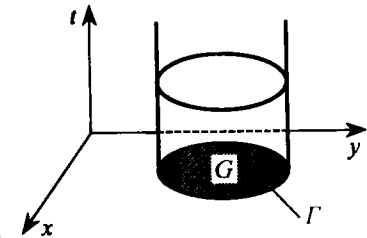


Рис. 4.7

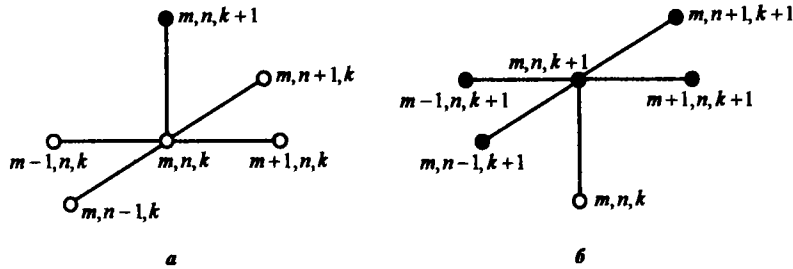


Рис. 4.8

Запишем разностную аппроксимацию уравнения (4.10) с использованием шеститочечных шаблонов двух типов, показанных на рис. 4.8.

На этом рисунке пустыми кружочками обозначены узлы на k -м слое с известными значениями функции, а зачерненными – узлы на $(k+1)$ -м слое, на котором параметры неизвестны и должны быть определены. Шаблон на рис. 4.8, а соответствует так называемой явной схеме, а шаблон на рис. 4.8, б – неявной схеме. В явной схеме оператор Лапласа аппроксимируется с использованием известных значений функции на k -м слое, а в неявной схеме – с использованием неизвестных значений функции на $(k+1)$ -м слое.

Без потерь общности будем далее полагать, что шаги по пространству одинаковы, т.е. $h_x = h_y = h$. Тогда явная разностная схема, соответствующая шаблону на рис. 4.8, а, запишется следующим образом:

$$\frac{u_{m,n}^{k+1} - u_{m,n}^k}{\tau} = \frac{u_{m+1,n}^k + u_{m-1,n}^k + u_{m,n+1}^k + u_{m,n-1}^k - 4u_{m,n}^k}{h^2} + O(\tau, h^2).$$

Из этого соотношения следует, что искомое значение определяется явным образом через известные значения на k -м слое по следующему соотношению:

$$u_{m,n}^{k+1} = u_{m,n}^k + \frac{\tau}{h^2} (u_{m+1,n}^k + u_{m-1,n}^k + u_{m,n+1}^k + u_{m,n-1}^k - 4u_{m,n}^k). \quad (4.13)$$

Для неявной разностной схемы, соответствующей шаблону на рис. 4.8, б, имеем

$$\frac{u_{m,n}^{k+1} - u_{m,n}^k}{\tau} = \frac{u_{m+1,n}^{k+1} + u_{m-1,n}^{k+1} + u_{m,n+1}^{k+1} + u_{m,n-1}^{k+1} - 4u_{m,n}^{k+1}}{h^2}. \quad (4.14)$$

Соотношение (4.14), записанное для всех внутренних узлов $(k+1)$ -го слоя, порождает систему линейных алгебраических уравнений для определения неизвестных значений функции в узлах. Каждое уравнение этой системы содержит только пять неизвестных $u_{m+1,n}^{k+1}, u_{m-1,n}^{k+1}, u_{m,n+1}^{k+1}, u_{m,n-1}^{k+1}, u_{m,n}^{k+1}$, хотя общее количество неизвестных равно числу внутренних узлов на $(k+1)$ -м слое и может быть весьма велико (число узлов $\approx N^2, N \approx 50 \div 100$). Следовательно, матрица

системы линейных уравнений является пятидиагональной и сильно разреженной. Для решения этой системы используются, как правило, итерационные методы, некоторые из которых описаны выше.

Очевидно, что при одинаковых шагах разностной сетки число операций, необходимых для отыскания решения на $(k+1)$ -м слое, в явных схемах значительно меньше, чем в неявных схемах. Однако в дальнейшем будет показано, что условие устойчивости разностной схемы накладывает значительные ограничения на величины шагов в явных схемах. Так как, в конечном счете, качество разностной схемы при одинаковой требуемой точности должно оцениваться количеством операций, необходимых для получения решения на всем временном интервале, то в ряде случаев неявные разностные схемы оказываются более предпочтительными, чем явные.

4.5. АППРОКСИМАЦИЯ, УСТОЙЧИВОСТЬ, СХОДИМОСТЬ РАЗНОСТНЫХ СХЕМ. ОСНОВНЫЕ ПОНЯТИЯ

Рассмотрим некоторую задачу, определяемую дифференциальным уравнением и граничными условиями. Запишем её в операторной форме:

$$Lu = f, \quad (4.15)$$

где L – некоторый дифференциальный оператор, действующий на искомую функцию u , f – правая часть. Примем, что оператор L включает как дифференциальные уравнения, так и граничные условия. На некоторой разностной сетке строим разностный оператор L_h , действующий на сеточную функцию u_h . Примером дифференциального оператора является оператор Лапласа Δ , а соответствующий ему разностный оператор возникает при аппроксимации вторых производных разностными ω отношениями (4.3) и (4.4).

При подстановке точного решения уравнения (4.15) в оператор L_h имеем

$$L_h u = f_h + \delta f_h,$$

где величина δf_h называется невязкой.

Определение 1. Разностная схема называется аппроксимирующей на решении, если норма невязки стремится к нулю при стремлении к нулю шага разностной схемы, т.е. $\|\delta f_h\| \rightarrow 0$ при $h \rightarrow 0$. Если при этом норма невязки удовлетворяет условию $\|\delta f_h\| \leq C_p h^p$ и константа C_p не зависит от h , то говорят, что разностная схема имеет p -й порядок аппроксимации или разностная схема аппроксимирует дифференциальный оператор с порядком p . Например, разностный оператор $L_h = \frac{u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1} - 4u_{m,n}}{h^2}$ аппроксимирует опера-

тор Лапласа $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ со вторым порядком точности.

Определение 2. Разностная схема называется сходящейся, если норма разности точного и приближенного решений $\|u - u_h\|$ стремится к нулю при стремлении к нулю шага разностной сетки. Если при этом $\|u - u_h\| \leq C_q h^q$, то говорят, что разностная схема имеет q -й порядок точности или имеет место сходимость с порядком q .

Порядок аппроксимации дифференциального уравнения не всегда совпадает с порядком точности разностной схемы, так как порядок точности разностной схемы зависит как от аппроксимации дифференциального уравнения, так и от аппроксимации граничных условий.

Не всякая аппроксимирующая схема является сходящейся. В связи с этим важным является понятие устойчивости разностной схемы

Определение 3. Устойчивостью разностной схемы называется непрерывная зависимость решения разностной задачи от правых частей и граничных условий. Для линейного оператора схема устойчива, если $\|u_n\| \leq C \|f_n\|$, где C – константа, не зависящая от шага разностной сетки и входных условий f .

Поясним понятие устойчивости. Ошибки при вычислении начальных и граничных условий и правых частей уравнений из-за ошибок округления и других причин можно рассматривать как возмущения начальных и граничных условий и правых частей уравнений. Очевидно, что разностная краевая задача (или задача с начальными данными) будет корректной и устойчивой, если решение разностной краевой задачи будет незначительно изменяться при малом изменении начальных и граничных условий и правых частей, связанном со случайными ошибками. В противном случае разностная краевая задача является неустойчивой. Важно отметить, что для неустойчивых разностных схем уменьшение шага сетки не приводит к устойчивости, поскольку любые малые возмущения решения со временем неограниченно возрастают.

Теорема. Пусть разностная задача $L_h u_h = f_h$ аппроксимирует оператор $Lu = f$ на решении с порядком p и устойчива. Тогда решение u_h сходится к точному решению и имеет место оценка $\|u - u_h\| \leq C_p h^p$, где C_p – константа, не зависящая от p и входных данных f .

4.6. ПРИМЕРЫ НЕУСТОЙЧИВЫХ РАЗНОСТНЫХ СХЕМ

Впервые с явлением неустойчивости при решении уравнений в частных производных столкнулись решая уравнение теплопроводности. Рассмотрим более простое уравнение – уравнение переноса, являющееся модельным для некоторых процессов в механике сплошной среды.

Рассмотрим задачу Коши для уравнения переноса

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} &= 0, \\ u(x, 0) &= u_0(x). \end{aligned} \quad (4.16)$$

Уравнение (4.16) является простейшим линейным уравнением гиперболического типа. Задача (4.16) имеет точное решение, которое строится с использованием метода характеристик. Точное решение дифференциального уравнения сравним, далее, с разностным решением, полученным с использованием различных схем, и дадим оценку аппроксимации, устойчивости и сходимости этих схем.

Уравнение (4.16) обладает семейством прямолинейных характеристик $x = t + \text{const}$, которые заполняют всю плоскость (t, x) (рис. 4.9). На каждой характеристике выполняется условие совместности $u = C$, где C – константа вдоль рассматриваемой характеристики, но меняющаяся от характеристики к характеристике. Отмеченные свойства позволяют легко построить решение задачи (4.16). Например, в некоторой точке A значение функции u_A равно значению функции u_0 в точке A' , расположенной на пересечении характеристики, выходящей из точки A , и оси абсцисс, на которой заданы данные Коши (рис. 4.9).

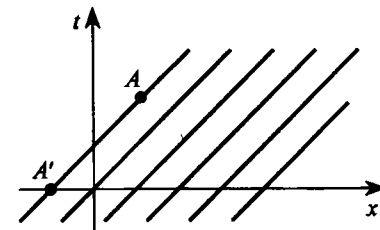


Рис. 4.9

Далее рассмотрим некоторую специальную начальную функцию вида

$$\begin{aligned} u_0(x, 0) &= u_0(x), \quad x < 0, \\ u_0(x, 0) &= 0, \quad x \geq 0. \end{aligned}$$

В силу свойств начальной функции правее характеристики AB (рис. 4.10), выходящей из начала координат, точное решение задачи таково, что $u = 0$. Левее этой характеристики точное решение определяется по значениям функции $u_0(x)$ указанным выше способом

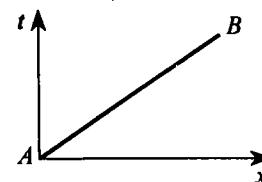


Рис. 4.10

Рассмотрим теперь две разностные схемы, с помощью которых попытаемся построить численное решение задачи (4.16). Трехточечные шаблоны этих схем изображены на рис. 4.11. Шаблон, изображенный на рис. 4.11, a , называют «уголок вперед» или «правый уголок», а шаблон, изображенный на рис. 4.11, b – «уголок назад» или «левый уголок».

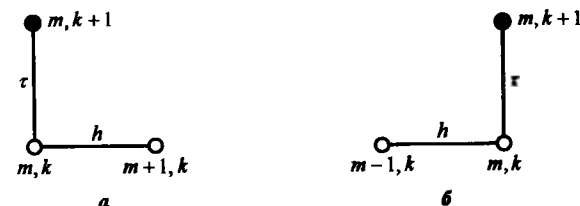


Рис. 4.11

Видно, что разностные схемы на этих шаблонах являются явными, поскольку для аппроксимации пространственной производной $\partial u/\partial x$ используются известные значения функции на k -м слое.

Для шаблона, показанного на рис. 4.11, а, разностная схема «уголок вперед» запишется в виде

$$\frac{u_m^{k+1} - u_m^k}{\tau} + \frac{u_{m+1}^k - u_m^k}{h} = 0. \quad (4.17)$$

Так как при аппроксимации производных использовались односторонние разности, то разностная схема (4.17) имеет первый порядок аппроксимации по τ и h . Из соотношения (4.17) решение на $(k+1)$ -м слое запишется в виде

$$u_m^{k+1} = u_m^k (1 + \tau/h) - u_{m+1}^k \tau/h. \quad (4.18)$$

Пусть $k=0$, тогда в силу граничных условий $u_m^0 = u_{m+1}^0 \equiv 0$ для $x \geq 0$ и из уравнения (4.18) следует, что при всех m и $x \geq 0$ решение на слое $k=1$ тождественно равно нулю, т.е. $u_m^1 \equiv 0$. Аналогично для всех последующих слоев $u_m^k \equiv 0$ при $x \geq 0$. Таким образом, из (4.18) следует, что в правом верхнем квадранте $u_h \equiv 0$, хотя, согласно точному решению, $u \equiv 0$ только в области, лежащей правее характеристики AB . Следовательно, разностная схема (4.17) дает решение, сколь угодно отличающееся от точного независимо от шагов разностной сетки τ и h ; т.е. при стремлении шагов τ и h к нулю не будет иметь места сходимость решения разностной задачи к точному решению. Вместе с тем разностная схема (4.17) является аппроксимирующей.

В рассмотренном случае, для схемы «уголок вперед», малое изменение граничных условий при $x < 0$ приводит к сколь угодно большим отличиям в решении, полученном по этой схеме в области между осью ординат и характеристикой, выходящей из начала координат. Таким образом, отсутствует непрерывная зависимость решения от граничных условий.

Для шаблона, показанного на рис. 4.11, б, разностная схема «уголок вперед» запишется в виде

$$\frac{u_m^{k+1} - u_m^k}{\tau} + \frac{u_m^k - u_{m-1}^k}{h} = 0. \quad (4.19)$$

Эта схема также имеет первый порядок аппроксимации. Решение на $(k+1)$ -м слое из соотношения (4.19) запишется следующим образом:

$$u_m^{k+1} = u_m^k (1 - \tau/h) + u_{m-1}^k \tau/h. \quad (4.20)$$

Рассмотрим три случая, соответствующие различным соотношениям шагов τ и h : $\tau/h = 1$, $\tau/h < 1$ и $\tau/h > 1$. В случае $\tau/h = 1$ (рис. 4.12, а) разностная схема дает точное решение $u_m^{k+1} = u_{m-1}^k$, поскольку значение функции по характеристике (помеченной на рисунке пунктиром) «переносится» из точки $(m-1, k)$ в точку $(m, k+1)$.

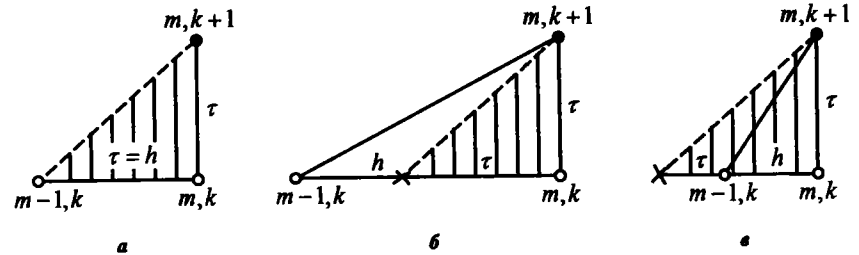


Рис. 4.12

Такое совпадение является, однако, свойством конкретных задач и разностной схемы, и в других случаях условие $\tau = h$ может и не обеспечивать точного решения.

В случае $\tau/h < 1$ (рис. 4.12, б) правая часть формулы (4.20) есть формула линейной интерполяции функции u_h между точками $(m-1, k)$ и (m, k) на слое k , которая дает значение функции в точке, отстоящей на расстояние τ от точки (m, k) . Это значение переносится по характеристике (пунктирная линия на рис. 4.12, б) в точку $(m, k+1)$ и дает решение, близкое к истинному с точностью, какой обладает линейная интерполяция. При этом, если $u_0(x)$ – линейная функция, то при $\tau/h < 1$ формула (4.20) дает точное решение.

При $\tau/h > 1$ (рис. 4.12, в) формула (4.20) есть формула линейной экстраполяции функции u_h и в общем случае можно получить решение, сильно отличающееся от точного.

Определение 4. Разностная схема называется абсолютно устойчивой, если она устойчива при любом соотношении шагов. Схема называется абсолютно неустойчивой, если она неустойчива при любом соотношении шагов. Разностная схема называется условно (частично) устойчивой, если она устойчива при некотором соотношении шагов.

Проведенное рассмотрение двух схем «уголок вперед» и «уголок назад» для уравнения переноса показывает, что схема (4.17) «уголок вперед» является абсолютно неустойчивой, а схема (4.19) «уголок назад» – условно устойчивой при $\tau/h \leq 1$. Условие $\tau/h < 1$ сформулируем в виде условия Куранта-Фридрикса-Леви (условие КФЛ).

Условие КФЛ. Пусть область влияния разностной схемы содержит в себе область влияния дифференциального уравнения. Разностная схема является неустойчивой, если это условие не выполнено.

На рис. 4.12 область влияния дифференциального уравнения заштрихована для всех трех случаев отношения τ/h . Видно, что для случая $\tau/h > 1$ разностная схема неустойчива.

4.7. ПРАКТИЧЕСКИЕ ПРАВИЛА ИССЛЕДОВАНИЯ УСТОЙЧИВОСТИ

Установить устойчивость разностной схемы с использованием данного выше определения на практике весьма затруднительно. Поэтому предложен ряд способов исследования устойчивости, позволяющих получить достаточные, а в ряде случаев необходимые и достаточные условия устойчивости разностных схем. Из методов исследования устойчивости упомянем спектральный признак устойчивости, метод гармоник Фурье, принцип максимума, метод операторных неравенств. Последний метод самый сильный из теоретических методов и дает строгие результаты для линейных уравнений с переменными (в том числе – разрывными) коэффициентами.

Исследование устойчивости разностных схем, аппроксимирующих сложные и в общем случае нелинейные уравнения, проводится следующим образом. Первоначально исходное дифференциальное уравнение заменяется более простым, модельным уравнением. Это достигается либо «замораживанием» коэффициентов в уравнениях с переменными коэффициентами, либо, в случае нелинейных уравнений, путем линеаризации последних. Для полученного модельного уравнения выписывается исследуемая разностная схема и рассматривается задача Коши при возмущенных граничных условиях и возмущенных правых частях. Исследуется эволюция решения во времени, и если возмущения с течением времени затухают, то разностная схема считается устойчивой, в противном случае – неустойчивой. Действительно, возмущения граничных условий и правых частей можно рассматривать как ошибки округления или аппроксимации, неизбежно возникающие при расчетах, и если разностная схема будет их усиливать, то решение будет неустойчивым.

4.8. СПЕКТРАЛЬНЫЙ ПРИЗНАК УСТОЙЧИВОСТИ

Пусть $u_m^{k+1} = S u_p^k$ ($p = i, i+1, i-1$). Оператор S , действующий на вектор u_p^k , называемый оператором перехода с одного временного слоя на другой, позволяет в результате последовательного применения трансформировать начальные данные в разностное решение на любом шаге по времени. Для линейных схем оператор S линеен. Так, для схемы «уголок назад» (4.19) имеем

$$u_m^{k+1} = u_m^k (1 + \tau/h) - u_{m-1}^k \tau/h = S[u_m^k, u_{m-1}^k]^T,$$

где S – матрица $\begin{vmatrix} 1 - \tau/h & 0 \\ 0 & \tau/h \end{vmatrix}$. Для линейных задач условие устойчивости по начальным данным разностной схемы есть условие

$$\|S\| \leq 1. \quad (4.21)$$

Действительно, $\|u^{k+1}\| = \|S u^k\| \leq \|S\| \cdot \|u^k\|$, т.е. $\|u^{k+1}\| \leq \|u^k\| \leq \dots \leq \|u^0\|$, и, следовательно, начальные возмущения затухают. Условие (4.21) есть условие

того, что спектр оператора S лежит внутри круга единичного радиуса на комплексной плоскости. Это означает, что $\max \|\lambda\| < 1$, где λ – собственное число оператора S . Достаточное условие устойчивости $|\lambda| < 1$. В связи с этим условие (4.21) называют спектральным признаком устойчивости. Более общее условие – условие Неймана: существует такая константа C , что $|\lambda| < 1 + \alpha(\tau)$.

4.9. ПРИНЦИП МАКСИМУМА

Перепишем уравнение (4.20) в виде $u_m^{k+1} = (1 - \gamma) u_m^k + \gamma u_{m-1}^k$, где $\gamma = \tau/h$. Так как $\gamma \leq 1$, то в этом соотношении коэффициенты при значениях функций в правой части больше нуля, и тогда соотношение (4.20) можно переписать в виде

$$\begin{aligned} |u_m^{k+1}| &= |(1 - \gamma) u_m^k + \gamma u_{m-1}^k| \leq [(1 - \gamma) + \gamma] \max_m (|u_m^k|, |u_{m-1}^k|) = \\ &= \max_m (|u_m^k|, |u_{m-1}^k|) \leq \max_m |u_m^k| \end{aligned}$$

Так как правая часть этого неравенства не зависит от m , то можно написать, что $\max_m |u_m^{k+1}| \leq \max_m |u_m^k|$. Применяя последовательно это неравенство, получаем $\max_m |u_m^{k+1}| \leq \max_m |u_m^k| \leq \dots \leq \max_m |u_m^0|$.

Последнее неравенство называют принципом максимума для разностной схемы: максимальное значение модуля решения разностной задачи достигается на границе области. Для задачи Коши, которую мы рассматривали, максимальное значение достигается при $t = 0$. Таким образом, начальные возмущения затухают.

Нетрудно видеть, что для схемы «уголок вперед» условие устойчивости по спектральному признаку и принципу максимума не будут выполнены

4.10. МЕТОД ГАРМОНИК ФУРЬЕ ИССЛЕДОВАНИЯ УСТОЙЧИВОСТИ РАЗНОСТНЫХ СХЕМ

Метод гармоник Фурье занимает одно из центральных мест среди методов исследования устойчивости разностных схем. Он используется для линейных разностных уравнений с постоянными коэффициентами на равномерной сетке. В методе гармоник Фурье при исследовании устойчивости разностных схем рассматривается задача Коши для линейного модельного уравнения для возмущений (добавок) начальных данных и правых частей. При этом рассматриваются возмущения (гармоники) специального вида.

Рассмотрим метод гармоник Фурье на примере задачи Коши для уравнения переноса

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} &= 0, \\ u(x, 0) &= u_0(x). \end{aligned} \quad (4.22)$$

Точное решение задачи можно записать в виде

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u_0(\theta) e^{i\theta(x-t)} d\theta,$$

где $i = \sqrt{-1}$. Решение задачи (4.22) представляется как суперпозиция бегущих монохроматических волн или гармоник вида $u_0(\theta) e^{i\theta(x-t)}$. Поскольку при разностной аппроксимации $t = k\tau$, $x = mh$, то соответствующую разностную гармонику можно представить в виде $u_m^k = \lambda^k e^{i\omega mh}$, а решение разностного уравнения

– в виде $u_m^k = \sum_{\omega=-\infty}^{+\infty} C_\omega \lambda^k e^{i\omega mh}$. Здесь λ – множитель, который выбирается таким

образом, чтобы u_m^k было решением разностной задачи, а ω – номер гармоники в разложении в ряд Фурье – определяется из разложения начальной функции (начального возмущения) в рассматриваемый ряд.

Множитель λ , называемый модулем перехода, в общем случае может быть комплексным числом и является показателем роста или затухания соответствующей гармоники. При $|\lambda| > 1$ начальные возмущения возрастают со временем, а при $|\lambda| < 1$ – затухают. Очевидно, что разностная схема будет устойчивой при условии $|\lambda| < 1$, поскольку любые возмущения правых частей и начальных условий (ошибки округления, аппроксимации и т.д.) будут затухать, так как $|e^{i\omega x}| < 1$.

Дж. фон Нейманом было сформулировано следующее условие устойчивости: $|\lambda| \leq 1 + m(\tau)$, которое также в ряде случаев является и необходимым, но не достаточным, условием сходимости. Ниже, однако, при изучении разностных схем будем использовать более жесткое условие: $|\lambda| \leq 1$.

4.11. ПРИМЕНЕНИЕ МЕТОДА ГАРМОНИК ФУРЬЕ ДЛЯ ИССЛЕДОВАНИЯ УСТОЙЧИВОСТИ РАЗНОСТНЫХ СХЕМ ДЛЯ УРАВНЕНИЯ ПЕРЕНОСА

Для уравнения (4.16) рассмотрим пять разностных схем, шаблоны которых показаны на рис. 4.13: явный «уголок вперед» (рис. 4.13, а), явный «уголок назад» (рис. 4.13, б), явная схема второго порядка точности (рис. 4.13, в), явная схема Лакса (рис. 4.13, г) и неявный «уголок назад» (рис. 4.13, д). Светлыми точками выделены узлы, в которых решение известно, а темными – узлы, в которых требуется определить решение.

Впишем некоторые вспомогательные формулы:

$$\left. \begin{aligned} e^{i\omega x} &= \cos(\omega x) + i \cdot \sin(\omega x), \\ 1 - \cos(\omega x) &= 2 \sin^2(\omega x/2), \\ \sin(\omega x) &= 2 \sin(\omega x/2) \cos(\omega x/2), \end{aligned} \right\} \quad (4.23)$$

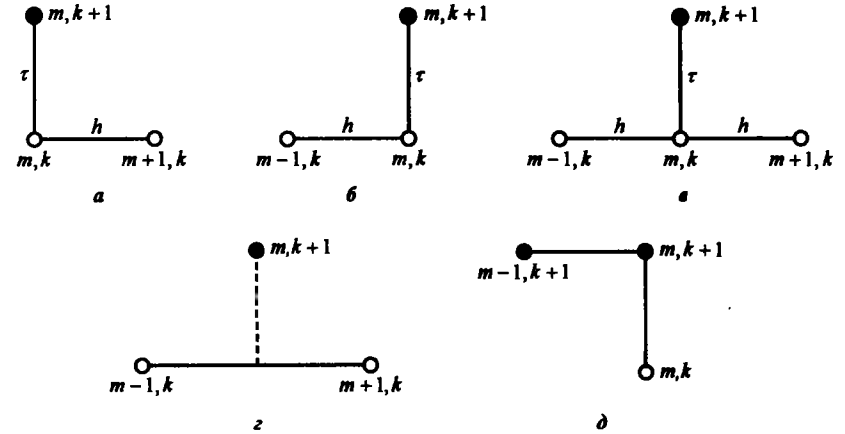


Рис. 4.13

$$\left. \begin{aligned} u_{m-1}^{(k+1)} &= \lambda^{k+1} \cdot e^{i\omega(x-h)}, \quad u_{m+1}^{(k+1)} = \lambda^{k+1} \cdot e^{i\omega(x+h)}, \\ u_m^{(k)} &= \lambda^k \cdot e^{i\omega x}, \quad u_{m-1}^{(k)} = \lambda^k \cdot e^{i\omega(x-h)}, \\ u_{m+1}^{(k)} &= \lambda^k \cdot e^{i\omega(x+h)}, \quad u_m^{(k+1)} = \lambda^{k+1} \cdot e^{i\omega x}. \end{aligned} \right\} \quad (4.24)$$

Для явного «уголка вперед» (рис. 4.13, а) имеем следующую разностную схему (см. формулу (4.17)):

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k)} - u_m^{(k)}}{h} = 0.$$

Используя (4.24), преобразуем (4.17) к виду

$$\frac{\lambda - 1}{\tau} + \frac{e^{i\omega h} - 1}{h} = 0$$

и, используя (4.23), – к виду

$$\lambda = 1 + [1 - \cos(\omega h)] \frac{\tau}{h} - i \sin(\omega h) \frac{\tau}{h}.$$

Для квадрата модуля λ имеем соотношение

$$|\lambda|^2 = \left[1 + 2 \sin^2(\omega h/2) \frac{\tau}{h} \right]^2 + \sin^2(\omega h) \frac{\tau^2}{h^2},$$

из которого следует, что $|\lambda| > 1$ при любых τ и h , т.е. разностная схема (4.17) – явный «уголок вперед» – абсолютно неустойчива. Этот же результат был получен выше из других рассуждений.

Для явного «уголка назад» (рис. 4.13, б) имеем следующую разностную схему (см. формулу (4.19)):

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_m^{(k)} - u_{m-1}^{(k)}}{h} = 0.$$

Выкладки, аналогичные проведенным для явного «уголка вперед», дают

$$|\lambda|^2 = 1 + 4 \sin^2(\omega h/2) \frac{\tau(\tau/h - 1)}{h}.$$

Из этого соотношения следует, что разностная схема (4.19) – явный «уголок назад», как это было показано и ранее, условно устойчива при $\tau/h \leq 1$.

Схемы (4.17) и (4.19) имеют первый порядок аппроксимации по τ и h , так как отброшены члены порядка малости $O(\tau h)$.

Рассмотрим *явную схему второго порядка точности* (рис. 4.13, в), в которой производная по x аппроксимируется со вторым порядком точности, т.е. отброшены члены малости $O(\tau, h^2)$. Разностная схема имеет вид

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k)} - u_{m-1}^{(k)}}{2h} = 0 \quad (4.25)$$

а для квадрата модуля λ может быть получено соотношение

$$|\lambda|^2 = 1 + \sin^2(\omega h) \frac{\tau^2}{h^2},$$

из которого следует, что разностная схема (4.25) абсолютно неустойчива, если использовать условие $|\lambda| < 1$. При условии $|\lambda| < 1 + O(\tau)$ схему можно рассматривать как частично устойчивую, если $\frac{\tau^2}{h^2} < C\tau$. Несмотря на повышение порядка аппроксимации по сравнению со схемой (4.17), при использовании условия $|\lambda| < 1$ схема абсолютно неустойчива.

Для *схемы Лакса* разностная схема имеет вид

$$\frac{u_m^{(k+1)} - [u_{m+1}^{(k)} + u_{m-1}^{(k)}]/2}{\tau} + \frac{u_{m+1}^{(k)} - u_{m-1}^{(k)}}{2h} = 0. \quad (4.26)$$

Схема Лакса имеет второй порядок точности по h и является условно устойчивой при $\tau/h \leq 1$. Эта схема обычно используется в задачах механики сплошных сред и принадлежит к схемам сквозного счета, так как позволяет без явного выделения проводить расчеты областей с поверхностями разрыва.

Для *схемы неявный «уголок назад»* (рис. 4.13, д) разностная схема имеет вид

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k+1)} - u_m^{(k+1)}}{h} = 0. \quad (4.27)$$

Для этой схемы $|\lambda| < 1$ независимо от соотношения шагов, и, следовательно, схема абсолютно устойчива.

4.12. ПРИМЕНЕНИЕ МЕТОДА ГАРМОНИК ФУРЬЕ В ЗАДАЧАХ ИССЛЕДОВАНИЯ УСТОЙЧИВОСТИ РАЗНОСТНЫХ СХЕМ ДЛЯ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ

Для одномерного уравнения теплопроводности

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad (4.28)$$

рассмотрим две разностные схемы – явную и неявную, шаблоны которых показаны соответственно на рис. 4.14, а и 4.14, б.

Для *явной схемы* имеем следующую разностную аппроксимацию с порядком $O(\tau, h^2)$:

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} = \frac{u_{m-1}^{(k)} - 2u_m^{(k)} + u_{m+1}^{(k)}}{h^2}. \quad (4.29)$$

Нетрудно получить, что модуль перехода λ для этой схемы равен

$$\lambda = 1 - \left(\frac{4\tau}{h^2}\right) \sin^2(\omega h/2).$$

Из этого соотношения следует, что явная схема (4.29) условно устойчива при $\frac{\tau}{h^2} \leq \frac{1}{2}$.

Для *неявной схемы* имеем следующую разностную аппроксимацию с порядком $O(\tau, h^2)$:

$$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} = \frac{u_{m-1}^{(k+1)} - 2u_m^{(k+1)} + u_{m+1}^{(k+1)}}{h^2}. \quad (4.30)$$

Можно показать, что для этой схемы $|\lambda| < 1$ при всех τ и h и, следовательно, схема абсолютно устойчива.

Полученные выше результаты исследования устойчивости различных разностных схем для уравнения переноса и одномерного уравнения теплопроводности сведены в табл. 4.2, при этом используется условие устойчивости $|\lambda| < 1$.

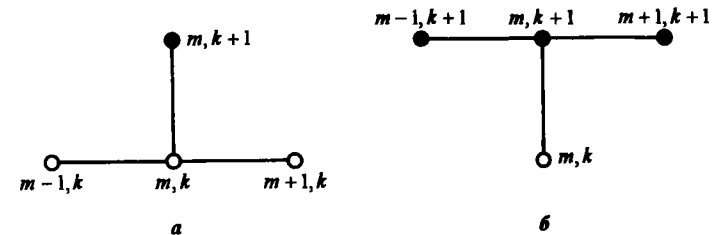


Рис. 4.14

Таблица 4.2

Дифференциальное уравнение	Шаблон	Условие устойчивости	Разностная схема	Порядок аппроксимации
$\frac{\partial u}{\partial t} \pm \frac{\partial u}{\partial x}$		Абсолютно неустойчива	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k)} - u_m^{(k)}}{h} = 0$	$O(\tau, h)$
		Условно устойчива $\tau/h \leq 1$	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_m^{(k)} - u_{m-1}^{(k)}}{h} = 0$	$O(\tau, h)$
		Абсолютно неустойчива	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k)} - u_{m-1}^{(k)}}{2h} = 0$	$O(\tau, h^2)$
		Условно устойчива $\tau/h \leq 1$	$\frac{u_m^{(k+1)} - [u_m^{(k)} - u_{m-1}^{(k)}]/2 + \frac{u_{m+1}^{(k)} - u_{m-1}^{(k)}}{2h}}{\tau} = 0$	$O\left(\tau, h^2, \frac{h^2}{\tau}\right)$
		Абсолютно устойчива	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} + \frac{u_{m+1}^{(k+1)} - u_m^{(k+1)}}{h} = 0$	$O(\tau, h)$
$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$		Условно устойчива $\tau/h^2 \leq 1/2$	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} = \frac{u_{m-1}^{(k)} - 2u_m^{(k)} + u_{m+1}^{(k)}}{h^2}$	$O(\tau, h^2)$
		Абсолютно устойчива	$\frac{u_m^{(k+1)} - u_m^{(k)}}{\tau} = \frac{u_{m-1}^{(k+1)} - 2u_m^{(k+1)} + u_{m+1}^{(k+1)}}{h^2}$	$O(\tau, h^2)$

Из представленных результатов следует, что неявные разностные схемы более устойчивы, чем явные разностные схемы, и, как правило, в неявных схемах не накладываются ограничения на шаги разностной сетки, связанные с устойчивостью. Однако при использовании неявных разностных схем для определения значений неизвестных на $(k+1)$ -м слое необходимо решать систему алгебраических уравнений относительно неизвестных $u_m^{(k+1)}$, $m = \overline{0, M}$. В то же время для явных схем значения неизвестных $u_m^{(k+1)}$ в каждой точке на $(k+1)$ -м слое явным образом выражаются через известные значения на k -м слое и их вычисление элементарно. Поэтому число операций для нахождения M значений неизвестных на одном слое в явных схемах значительно меньше, чем в неявных схемах. В то же время из-за ограничений, связанных с устойчивостью разностных схем, число шагов по времени для достижения конечного значения T может быть значительно больше в явных схемах, чем в неявных.

Таким образом, предпочтительность той или иной разностной схемы при заданной точности должна определяться с учетом общего числа операций, необходимого для полного цикла вычислений. С этой точки зрения существуют ситуации, когда может быть предпочтительнее та или иная схема – явная или неявная. Например, для одномерных по пространству задач часто более выгодными оказываются неявные разностные схемы.

4.13. МЕТОД ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ (МЕТОД ДРОБНЫХ ШАГОВ)

На практике приходится иметь дело с многомерными задачами, когда необходимо учитывать изменение искомых функций по координатам x, y, z . Для таких задач выбор разностной схемы является нетривиальным и требует анализа различных аспектов. Можно показать, что расчет по явной схеме до заданного момента времени T с учетом ограничений на шаги, связанных с устойчивостью, требует по явной схеме $\sim N^{p+2}$ операций, а по неявной схеме $\sim N^{3p-1}$ операций. Здесь N – число точек на слое, а p – пространственная размерность задачи. Для рассматриваемого здесь метода переменных направлений требуется N^{p+1} операций. В табл. 4.3 представлено число операций для различных схем и различных пространственных размерностей задачи. Из таблицы следует, что для задач с пространственной размерностью больше единицы более предпочтительным является метод переменных направлений.

Таблица 4.3

	Явная схема	Неявная схема	Метод переменных направлений
1	N^3	N^2	N^2
2	N^4	N^5	N^3
3	N^5	N^8	N^4

Метод переменных направлений рассмотрим на примере решения двумерного уравнения теплопроводности – уравнения параболического типа с двумя независимыми пространственными переменными

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f(x, y, t). \quad (4.31)$$

Требуется найти решение этого уравнения внутри параллелепипеда $[0 \leq x \leq a, 0 \leq y \leq b, 0 \leq t \leq T]$ при начальном условии

$$u|_{t=0} = u_0(x, y, 0) \quad (4.32)$$

и граничных условиях на гранях параллелепипеда

$$u|_{\Gamma} = u^0(x, y, t). \quad (4.33)$$

Идея метода переменных направлений состоит в сведении решения трехмерной задачи по переменным t, x и y к решению последовательности двумерных задач по переменным t, x и y .

Построим разностную схему. Пусть τ, h_x и h_y – шаги соответственно по t, x и y и $x_m = mh_x, y_n = nh_y, t_k = k\tau, m = \overline{0, M}, n = \overline{0, N}, k = \overline{0, K}$. Введем разностные операторы

$$L_x[u_{m,n}^{(k)}] = -\frac{u_{m-1,n}^{(k)} - 2u_{m,n}^{(k)} + u_{m+1,n}^{(k)}}{h_x^2}, \quad m = \overline{1, M-1},$$

$$L_y[u_{m,n}^{(k)}] = -\frac{u_{m,n-1}^{(k)} - 2u_{m,n}^{(k)} + u_{m,n+1}^{(k)}}{h_y^2}, \quad n = \overline{1, N-1},$$

$$L[u_{m,n}^{(k)}] = L_x[u_{m,n}^{(k)}] + L_y[u_{m,n}^{(k)}], \quad k = \overline{1, K-1}.$$

Аппроксимируя уравнение (4.31) разностным, можем записать

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k)}}{\tau} + (L_x + L_y)[\sigma u_{m,n}^{(k+1)} + (1-\sigma)u_{m,n}^{(k)}] = f_{m,n},$$

где σ – параметр разностной схемы, который может принимать значения от 0 до 1. Погрешность аппроксимации такой схемы $O(\tau^v, h_x^2, h_y^2)$, где $v=2$ при $\sigma=1/2$ и $v=1$ при $\sigma \neq 1/2$. Для явной схемы $\sigma=0$, для чисто неявной схемы $\sigma=1$.

Рассмотрим шаблоны, изображенные на рис. 4.15. Отметим, что шаблон 4.15, в не переносится на трехмерный случай.

Явная разностная схема (рис. 4.15, а)

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k)}}{\tau} + L[u_{m,n}^{(k)}] = f_{m,n} \quad (4.34)$$

устойчива при $\tau/h^2 \leq 1/4$ (полагаем $h = h_x = h_y$). Схема позволяет по известным значениям функции на временном слое k найти решение на $(k+1)$ -м слое.

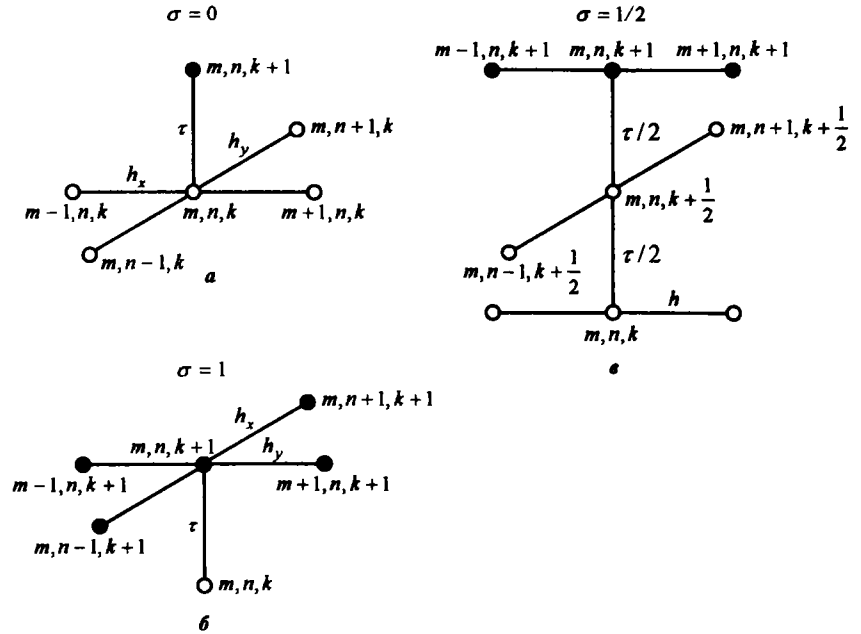


Рис. 4.15

Неявная разностная схема (рис. 4.15, б)

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k)}}{\tau} + L[\sigma u_{m,n}^{(k+1)} + (1-\sigma)u_{m,n}^{(k)}] = f_{m,n} \quad (4.35)$$

абсолютно устойчива и позволяет определить решение на $(k+1)$ -м слое путем решения системы линейных алгебраических уравнений с ленточной матрицей с широкой слабо заполненной лентой. Как следует из табл. 4.3, явная и неявная схемы требуют при $P=2 \sim N^4$ и $\sim N^5$ операций соответственно, но и в том и в другом случае требуемое количество операций весьма велико. Между тем метод переменных направлений требует лишь N^2 операций.

Рассмотрим разностную схему метода переменных направлений. Для этого введем дополнительный промежуточный слой $(k+1/2)$ и следующую разностную аппроксимацию уравнения (4.31) на этом слое на шаблоне (рис. 4.15, в):

$$\frac{u_{m,n}^{(k+1/2)} - u_{m,n}^{(k)}}{\tau/2} + L_x(u_{m,n}^{(k)}) + L_y(u_{m,n}^{(k+1/2)}) = f_{m,n}^{(k)}. \quad (4.36)$$

Согласно (4.36) вторая производная по x аппроксимируется на k -м слое, а по y – на $(k+1/2)$ -м слое. Поэтому на $(k+1/2)$ -м слое имеем только три не-

известных значения $u_{m,n-1}^{(k+1/2)}$, $u_{m,n}^{(k+1/2)}$, $u_{m,n+1}^{(k+1/2)}$ в каждом уравнении системы, хотя требуется найти $(N-1)^2$ неизвестных (полагая $N=M$).

Итак, во-первых, на $(k+1/2)$ -м слое имеем систему линейных уравнений с трехдиагональной матрицей, которую можно записать в виде

$$\frac{\tau}{2h_y^2} u_{m,n-1}^{(k+1/2)} - \left(1 + \frac{\tau}{2h_y^2}\right) u_{m,n}^{(k+1/2)} + \frac{\tau}{2h_y^2} u_{m,n+1}^{(k+1/2)} = F_{m,n}^k, \quad (4.37)$$

где $F_{m,n}^k = u_{m,n}^{(k)} - \frac{\tau}{2} (L_x[u_{m,n}^{(k)}] - f_{m,n}^{(k)})$ вычисляется по известным значениям функции на k -м слое. Присоединив к (4.37) значения $u_{m,0}^{(k+1/2)}$ и $u_{m,N}^{(k+1/2)}$, известные из краевых условий, можем вычислить значения неизвестных на $(k+1/2)$ -м слое. Для этого путем решения системы линейных уравнений с трехдиагональной матрицей можно использовать такой эффективный метод решения, как метод прогонки (см. гл. 1).

Во-вторых, система линейных уравнений (4.37) распадается на $M-1$ независимых трехдиагональных систем порядка $M-1$ для каждого m . Независимое решение на $(k+1/2)$ -м слое при фиксированном m позволяет существенно уменьшить объем вычислений по сравнению с решением одной системы порядка $(M-1)^2$.

После того, как найдены все неизвестные $u_{m,n}^{(k+1/2)}$ на промежуточном слое, решение на $(k+1)$ -м слое определяется следующей разностной аппроксимацией уравнения (4.31):

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k+1/2)}}{\tau/2} + L_x(u_{m,n}^{(k+1)}) + L_y(u_{m,n}^{(k+1/2)}) = f_{m,n}^{(k+1/2)}. \quad (4.38)$$

Соотношение (4.38) может быть переписано в виде

$$\frac{\tau}{2h_x^2} u_{m-1,n}^{(k+1)} - \left(1 - \frac{\tau}{2h_x^2}\right) u_{m,n}^{(k+1)} + \frac{\tau}{2h_x^2} u_{m+1,n}^{(k+1)} = F_{m,n}^{(k+1/2)}, \quad (4.39)$$

где $F_{m,n}^{(k+1/2)} = -\frac{\tau}{2} (L_y[u_{m,n}^{(k+1/2)}] - f_{m,n}^{(k+1/2)}) + u_{m,n}^{(k+1/2)}$ вычисляется по известным значениям функции на $(k+1/2)$ -м слое. Аналогично ситуации на предыдущем промежуточном слое система (4.39) (с присоединенными граничными значениями) распределяется на $(N-1)$ независимых трехдиагональных линейных систем, отвечающих фиксированным значениям n , каждая из которых решается методом прогонки. Только прогонка теперь выполняется по индексу m , т.е. в направлении оси x . Отметим, что коэффициенты при неизвестных в уравнениях (4.37) и (4.39) таковы, что прогонка является корректной и устойчивой.

На рис. 4.16 схематически изображен переход со слоя k на слой $(k+1)$ с дробным шагом и указано направление прогонок (прямой и обратный ход).

Таким образом, в методе переменных направлений переход с известного k -го слоя на искомый $(k+1)$ -й слой происходит с использованием дробного шага $\tau/2$ и промежуточного слоя $(k+1/2)$. Разностная схема (4.36) и (4.38) в совокупности является абсолютно устойчивой, хотя при переходе со слоя k на слой $(k+1/2)$ модуль перехода $|\lambda_x| > 1$, поскольку схема (4.39) как бы локально неустойчива. Однако при переходе со слоя $(k+1/2)$ на слой k модуль перехода $|\lambda_y| < 1$ и общее значение модуля перехода $|\lambda| = |\lambda_x| \cdot |\lambda_y|$ оказывается меньше единицы.

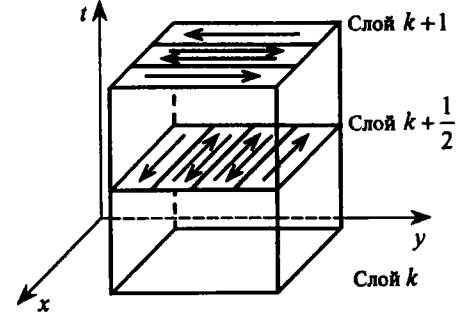


Рис. 4.16

Однако при переходе со слоя $(k+1/2)$ на слой k модуль перехода $|\lambda_y| < 1$ и общее значение модуля перехода $|\lambda| = |\lambda_x| \cdot |\lambda_y|$ оказывается меньше единицы.

Отметим, что существуют так называемые локально-одномерные схемы (или схемы расщепления), в которых на промежуточном $(k+1/2)$ -м слое отсутствуют и аппроксимация, и устойчивость. Тем не менее при переходе со слоя k на слой $(k+1)$ результирующая схема становится аппроксимирующей и устойчивой. Например, для уравнения (4.31) при $f=0$ предложена локально-одномерная разностная схема: при переходе от слоя k к слою $(k+1/2)$

$$\frac{u_{m,n}^{(k+1/2)} - u_{m,n}^{(k)}}{\tau} = L_y[u_{m,n}^{(k+1/2)}] \quad (4.40)$$

с порядком аппроксимации $O(\tau, h^2)$ и при переходе со слоя $(k+1/2)$ на слой $(k+1)$

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k+1/2)}}{\tau} = L_x[u_{m,n}^{(k+1)}] \quad (4.41)$$

с порядком аппроксимации $O(\tau, h^2)$. Видно, что на промежуточных слоях ни одна из схем (4.40) и (4.41) не является аппроксимирующей, хотя результирующая схема — аппроксимирующая и устойчивая с порядком аппроксимации $O(\tau, h^2)$.

4.14. МЕТОД УСТАНОВЛЕНИЯ – МАТЕМАТИЧЕСКИЕ И ФИЗИЧЕСКИЕ ОСНОВЫ МЕТОДА

Метод установления обычно используется для решения уравнений эллиптического типа или смешанных эллиптико-гиперболических уравнений. В задачах газовой динамики уравнения эллиптического типа описывают стационарное

дозвуковое течение, например течение за отошедшей ударной волной при сверхзвуковом обтекании затупленных тел, или течение в сужающейся части сопла Лавала. Уравнение для функции тока и потенциала скорости в плоском несжимаемом течении является уравнением Лапласа. К эллиптическому уравнению приводит задача определения напряжений, возникающих при упругом кручении длинного цилиндрического стержня, задача о стационарных потоках тепла в двумерном теле и многие другие.

Для уравнений эллиптического типа характерной является задача Дирихле, когда на границе Γ области G , в которой ищется решение, задана искомая функция $u(x, y)$ (рис. 4.17, а). В связи с этим идею метода установления рассмотрим на примере задачи Дирихле для уравнения Лапласа с двумя независимыми переменными x и y :

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad (4.42)$$

$$u(x, y)|_{\Gamma} = u^0(x, y).$$

Задачу (4.42) будем называть стационарной, так как решение задачи не зависит от времени. Наряду со стационарной задачей рассмотрим эволюционную (нестационарную) задачу для параболического уравнения теплопроводности с теми же граничными условиями и произвольно выбранными начальными данными (рис. 4.17, б):

$$\frac{\partial u}{\partial t} = \Delta u,$$

$$u(x, y, t)|_{\Gamma} = u^0(x, y), \quad (4.43)$$

$$u(x, y, 0) = u_0(x, y).$$

В общем случае задачу (4.43) называют нестационарной краевой задачей, если граничные условия зависят от времени. В представленной формулировке задача (4.43) называется вырожденной нестационарной задачей, так как условие на границе области не зависит от времени.

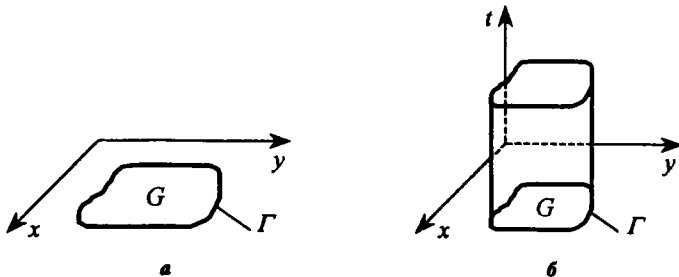


Рис. 4.17

Таким образом, в методе установления вводится новая независимая переменная t и задача формально усложняется – увеличивается размерность задачи. Однако при этом увеличивается набор возможных разностных схем решения задачи, что допускает выбор из них устойчивых и экономичных. Ниже будет показано, что кажущееся усложнение задачи не приводит ни к увеличению времени счета, ни к потере точности. Важным достоинством метода установления является то, что краевая задача для эллиптического уравнения сводится к задаче с начальными и краевыми условиями. Из физических соображений достаточно очевидно, что решение эволюционной задачи (4.43) при $t \rightarrow \infty$ будет стремиться к решению стационарной задачи (4.42).

Действительно, решение стационарной задачи (4.42) дает стационарное распределение значений искомой функции (например, температуры) в области G при заданном её распределении на границе области Γ . Решение эволюционной задачи (4.43) дает распределение значений функции во времени при тех же, что и в стационарной задаче, граничных условиях. Очевидно, что по истечении достаточно большого промежутка времени в области G установится распределение значений функции, не зависящее от начальных условий, а определяемое лишь значениями функции на границе. Приведенное рассуждение дает основание предполагать, что решение эволюционной задачи при $t \rightarrow \infty$ сводится к решению стационарной задачи и решение эволюционной задачи при достаточно большом $t = T$ может быть принято за приближенное решение исходной стационарной задачи.

Приведем некоторые математические рассуждения, подтверждающие сделанное предположение о выходе решения эволюционной задачи (4.43) на стационарный режим. Рассмотрим разность

$$w(x, y, t) = v(x, y, t) - u(x, y),$$

где (x, y, t) и $u(x, y)$ – решение задач (4.43) и (4.42) соответственно.

Функция w удовлетворяет дифференциальному уравнению (4.43). Начальными данными являются

$$w(x, y, 0) = v_0(x, y) - u(x, y),$$

а граничные условия – однородные:

$$w(x, y, t) = 0.$$

Используя метод разделения переменных Фурье для получения решения линейной однородной краевой задачи для функции w , покажем, что $w(x, y, t) \rightarrow 0$ при $t \rightarrow \infty$. Рассмотрение проведем на примере соответствующей одномерной задачи в области $G = [0 \leq t \leq T, 0 \leq x \leq 1]$. Имеем:

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2}, \quad (4.44)$$

$$w(0, t) = w(1, t) = 0, \quad (4.45)$$

$$w(x,0) = u_0(x) - u(x). \quad (4.46)$$

Будем искать частное решение задачи в виде

$$w(x,t) = X(x) Y(t). \quad (4.47)$$

Подставив (4.47) в (4.44), получаем

$$\frac{X''(x)}{X(x)} = \frac{Y'(t)}{Y(t)} = -\lambda, \quad (4.48)$$

где $\lambda = \text{const}$. Подставляя (4.47) в (4.45), получаем

$$\begin{aligned} w(0,t) &= X(0) Y(t) = 0, \\ w(1,t) &= X(1) Y(t) = 0. \end{aligned} \quad (4.49)$$

Из этих соотношений следует, что $X(0) = X(1) = 0$. Из (4.48) получаем два обыкновенных дифференциальных уравнения для нахождения функций $X(x)$ и $Y(t)$:

$$\frac{dY}{dt} + \lambda Y = 0, \quad (4.50)$$

$$\frac{d^2 X}{dx^2} + \lambda X = 0. \quad (4.51)$$

Для уравнения (4.51) решаем краевую задачу с однородными условиями (4.49). Имеем простейшую задачу на собственные значения, из которой необходимо определить значения параметра λ , при которых существуют нетривиальные решения задачи (4.51), (4.49).

Рассмотрим три возможных случая:

1. $\lambda < 0$; $X(x) = C_1 e^{\sqrt{-\lambda}x} + C_2 e^{-\sqrt{-\lambda}x}$. Из (4.49) имеем $C_1 = C_2 = 0$; $X(x) = 0$;
2. $\lambda = 0$; $X(x) = C_1 x + C_2$. Из (4.49) вновь имеем $C_1 = C_2 = 0$; $X(x) = 0$;
3. $\lambda > 0$; $X(x) = C_1 \cos(\sqrt{\lambda}x) + C_2 \sin(\sqrt{\lambda}x)$. Подставляя в (4.49), получаем $X(x) = 0$; $C_1 = 0$; и $X(1) = C_2 \sin \sqrt{\lambda} = 0$ и, если $X(x) \neq 0$, то $C_2 \neq 0$.

Нетривиальные решения задачи (4.51), (4.49) возможны при

$$\lambda = \lambda_j = (\pi j)^2, \quad j = 1, 2, \dots$$

Интегрируя (4.50), получаем $Y(t) = C_j e^{-\lambda_j t}$. Согласно (4.47) частное решение $w(x,t)$ имеет вид

$$w(x,t) = C_j e^{-\lambda_j t} \sin(\pi j x).$$

Константы C_j определяются из разложения начальной функции $w(x,0)$ в ряд Фурье. Нам сейчас не интересуют конкретные формулы для вычисления этих констант. Из последнего соотношения следует, что $w(x,t) \rightarrow 0$ при $t \rightarrow \infty$, т.е. решение эволюционной задачи стремится при $t \rightarrow \infty$ к решению стационарной задачи. Таким образом, решение задачи (4.42) можно рассматривать как предел решения нестационарной задачи (4.43) при $t \rightarrow \infty$.

Рассмотрим еще одну эволюционную задачу:

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} &= \Delta u, \quad u|_{\Gamma} = u^0(x, y), \\ u|_{t=0} &= u_0(x, y), \quad \frac{\partial u}{\partial t}|_{t=0} = u_1(x, y), \end{aligned}$$

описывающую колебания мембраны при наличии вязкого трения. Колебания со временем будут затухать, и решение этой задачи будет стремиться к решению задачи (4.42), которую можно рассматривать как задачу о прогибе мембраны.

Если в уравнении отбросить член $\frac{\partial u}{\partial t}$, то оно будет описывать процесс колебания мембраны без затухания амплитуды колебаний и решение устанавливаться не будет.

4.15. РАЗНОСТНЫЕ СХЕМЫ МЕТОДА УСТАНОВЛЕНИЯ

Для простоты будем рассматривать прямоугольную область $G = [0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq t \leq T]$. Простейшей разностной аппроксимацией стационарной задачи (4.42) является следующая:

$$\frac{u_{m,n+1} - 2u_{m,n} + u_{m-1,n}}{h_x^2} + \frac{u_{m,n+1} - 2u_{m,n} + u_{m,n-1}}{h_y^2} = 0, \quad (4.52)$$

где $x_m = mh$, $y_n = nh$, $u_{m,n} = u(x_m, y_n)$. Для решения этой системы линейных уравнений с пятидиагональной матрицей может быть использован один из способов рассмотренных ранее итерационных процессов.

Для аппроксимации задачи (4.43) введем сетку:

$$\begin{aligned} x_m &= mh, \quad y_n = nh, \quad t_k = k\tau, \\ h_x &= 1/M, \quad h_y = 1/N, \quad m = \overline{0, M}, \quad n = \overline{0, N}, \quad k = \overline{0, K} \end{aligned}$$

и обозначение $u_{m,n}^{(k)} = u(x_m, y_n, t_k)$. Для дифференциального уравнения (4.43) запишем явную и неявную разностные схемы, шаблоны которых показаны соответственно на рис. 4.18, а и б.

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k)}}{\tau} = \frac{u_{m+1,n}^{(k)} - 2u_{m,n}^{(k)} + u_{m-1,n}^{(k)}}{h_x^2} + \frac{u_{m,n+1}^{(k)} - 2u_{m,n}^{(k)} + u_{m,n-1}^{(k)}}{h_y^2}, \quad (4.53)$$

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k)}}{\tau} = \frac{u_{m+1,n}^{(k+1)} - 2u_{m,n}^{(k+1)} + u_{m-1,n}^{(k+1)}}{h_x^2} + \frac{u_{m,n+1}^{(k+1)} - 2u_{m,n}^{(k+1)} + u_{m,n-1}^{(k+1)}}{h_y^2}. \quad (4.54)$$

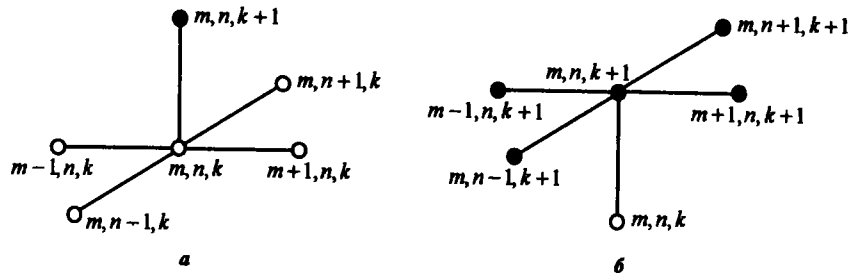


Рис. 4.18

Большинство итерационных методов (в том числе метод простой итерации и метод Зейделя) можно символически записать в виде

$$B_k \frac{u^{(k+1)} - u^{(k)}}{\tau_k} + Au^{(k)} = f.$$

При этом, если B_k и τ_k не зависят от номера итерации k , то процесс называется стационарным. В связи с этим разностные схемы (4.53) и (4.54) решения эволюционной задачи можно интерпретировать как итерационный метод решения системы линейных алгебраических уравнений, возникающей при разностной аппроксимации стационарного уравнения (4.42). При этом для схемы (4.53) $B = E$ и $\tau_k = \tau$ (здесь E – единичная матрица). С этой точки зрения можно оценивать достоинства и недостатки метода установления. Если для достижения заданной точности в методе установления потребуется меньшее число итераций (т.е. шагов по времени), чем при решении системы уравнений (4.52) каким-либо другим итерационным методом, то использование метода установления целесообразно.

Таким образом, при использовании метода установления важно понимать, что итерационный процесс можно рассматривать как разностную схему решения некоторой вырожденной нестационарной (эволюционной) задачи.

Исследование устойчивости для новой разностной схемы (4.53) для возмущения специального вида

$$u_{m,n}^{(k)} = \lambda^k e^{i\omega_x m h_x} e^{i\omega_y n h_y}, \quad (4.55)$$

(где i – мнимая единица) дает для модуля перехода следующее соотношение:

$$\lambda = 1 - \left(\frac{4\tau}{h_x^2} \right) \sin^2(\omega_x h_x / 2) - \left(\frac{4\tau}{h_y^2} \right) \sin^2(\omega_y h_y / 2). \quad (4.56)$$

Отсюда, учитывая, что ω_x и ω_y могут принимать произвольные значения, находим условие устойчивости: $\tau \left(\frac{1}{h_x^2} + \frac{1}{h_y^2} \right) \leq \frac{1}{2}$, или в частном случае квадратной сетки, когда $h_x = h_y = h$:

$$\frac{\tau}{h^2} \leq \frac{1}{4}. \quad (4.57)$$

Для неявной разностной схемы (4.54) модуль перехода определяется соотношением

$$\lambda = \left[1 + \left(\frac{4\tau}{h_x^2} \right) \sin^2(\omega_x h_x / 2) + \left(\frac{4\tau}{h_y^2} \right) \sin^2(\omega_y h_y / 2) \right]^{-1}, \quad (4.58)$$

из которого следует, что неявная разностная схема (4.54) абсолютно устойчива, так как $|\lambda| < 1$ при любых шагах τ , h_x и h_y .

Условие устойчивости (4.57) является весьма жестким. Оно, как правило, накладывает более сильные ограничения на шаг, чем естественные требования точности. В случае двух (и более) пространственных переменных применение неявных схем вызывает большие трудности, связанные с решением системы линейных уравнений. Это обстоятельство послужило одним из стимулов к развитию родственных методов, получивших различные названия: метод переменных направлений, метод дробных шагов, метод расщепления и др.

Приведем пример, дающий некоторое представление о существе дела. Рассмотрим квадратную сетку с шагом $h_x = h_y = h$. Введем полуцелые временные слои для аппроксимации уравнения (4.43). Переход со слоя k на слой $k+1$ разобьем на два «полушага»:

$$\frac{u_{m,n}^{(k+1/2)} - u_{m,n}^{(k)}}{\tau} = \frac{u_{m+1,n}^{(k+1/2)} - 2u_{m,n}^{(k+1/2)} + u_{m-1,n}^{(k+1/2)}}{h^2}, \quad (4.59)$$

$$\frac{u_{m,n}^{(k+1)} - u_{m,n}^{(k+1/2)}}{\tau} = \frac{u_{m,n+1}^{(k+1/2)} - 2u_{m,n}^{(k+1/2)} + u_{m,n-1}^{(k+1/2)}}{h^2}. \quad (4.60)$$

Отметим, что схемы (4.59) и (4.60) – неявные и содержат неизвестные значения на слое $k+1/2$ в схеме (4.59) и на слое $k+1$ в схеме (4.60). При этом в схеме (4.59) неизвестные значения функции вычисляются прогонкой по слоям $y = \text{const}$, в схеме (4.60) – прогонкой по слоям $x = \text{const}$. Схема (4.59), (4.60) получила свое название – метод расщепления, метод дробных шагов, метод переменных направлений – из-за расщепления уравнения (4.43) на два уравнения, введения дробного шага и осуществления прогонки попеременно – сначала по направлению y , а затем по направлению x .

Естественно возникает вопрос о том, в каком смысле пара уравнений (4.59) и (4.60) приближает исходное уравнение (4.43). Заметим, что данное ранее определение аппроксимации здесь явно не годится: сеточные уравнения (4.59) и (4.60) не аппроксимируют уравнение (4.43) на дробных шагах. Однако можно показать, что для перехода со слоя k на слой $k+1$ результирующая схема является аппроксимирующей. Кроме того, если исходить из физических

соображений, то имеются очень веские причины надеяться, что совокупность сеточных уравнений (4.59) и (4.60) приближенно описывает процесс распространения тепла. Действительно, уравнение (4.59) представляет сеточную аппроксимацию предельно анизотропного процесса теплопередачи, при котором распространение тепла происходит лишь в направлении оси x ; аналогично можно истолковать уравнение (4.60). Можно надеяться, что при соответствующем подборе коэффициентов теплопроводности такое попеременное распределение тепла по двум направлениям будет приближенно описывать реальный (изотропный) процесс теплопроводности.

Покажем, что модуль перехода, соответствующий двум «полушагам», действительно аппроксимирует модуль перехода для уравнения (4.43). Методом гармоник Фурье можем получить выражения для модулей перехода разностных схем (4.59) и (4.60):

$$\lambda' = \left[1 + \left(\frac{4\tau}{h^2} \right) \sin^2(\omega_1 h/2) \right]^{-1},$$

$$\lambda'' = \left[1 + \left(\frac{4\tau}{h^2} \right) \sin^2(\omega_2 h/2) \right]^{-1}.$$

Для совокупности переходов (4.59) и (4.60) модуль перехода $\lambda = \lambda' \cdot \lambda''$. Из полученного соотношения для модуля перехода очевидно, что схема (4.59), (4.60) абсолютно устойчива, т.е. устойчива при любых значениях шагов τ и h .

При $h \rightarrow 0$ можно записать приближенное соотношение для модуля перехода: $\lambda \approx 1 - \left(\frac{4\tau}{h^2} \right) \sin^2(\omega_1 h/2) - \left(\frac{4\tau}{h^2} \right) \sin^2(\omega_2 h/2) \approx 1 - \tau(\omega_1^2 + \omega_2^2)$.

С другой стороны, для точного решения уравнения теплопроводности $u = e^{-(\omega_1^2 + \omega_2^2)\tau} e^{i\omega_1 x} e^{i\omega_2 y}$ модуль перехода определяется соотношением $\lambda = e^{-(\omega_1^2 + \omega_2^2)\tau} \approx 1 - \tau(\omega_1^2 + \omega_2^2)$.

Таким образом, имеет место аппроксимация по меньшей мере в следующем смысле: при переходе со слоя k на слой $k+1$ решение специального вида для сеточных уравнений изменяется так же, как и соответствующее решение дифференциального уравнения (4.43), с точностью до малых более высокого порядка, чем τ . Используя преобразование Фурье, можно строго обосновать сходимость приближенного решения к точному.

Для возмущения специального вида (4.55) скорость сходимости определяется модулем перехода λ , который для явной схемы (4.53) имеет вид (4.56). Наиболее медленно убывают низкочастотные возмущения, для которых $\lambda \approx 1 - \tau(\omega_1^2 + \omega_2^2)$. Условие устойчивости (4.57) ограничивает сверху допустимый шаг τ ; положим $\tau = \frac{1}{4} \alpha h^2$, $0 < \alpha < 1$. Заметим далее, что минимальные значения частот ω_1 и ω_2 определяются размером области; в нашем случае

$\omega_{1,2} \approx \pi$. Таким образом, для минимальных частот имеем $\lambda \approx 1 - \frac{\pi^2}{4} \alpha h^2$. Из этого соотношения следует, что при малых значениях h сходимость весьма медленная. Для того чтобы уменьшить в $1/\varepsilon$ раз амплитуду ошибки, требуется $N = N(\varepsilon)$ слоев, где N определяется соотношениями $|\lambda|^N \approx \varepsilon^{-1}$, $N \approx \frac{|\ln \varepsilon|}{|\ln \lambda|} \approx |\ln \varepsilon| \frac{4}{\alpha \pi^2} \frac{1}{h^2}$.

Например, для $\varepsilon = 0.01$, $\alpha = 1$ и $h = 0.01$ имеем $N = 8000$. Из этого примера видно, что простейшие итерационные методы, основанные на явных аппроксимациях нестационарной краевой задачи, могут оказаться неэффективными.

Рассмотрим теперь неявную аппроксимацию (4.59), (4.60), построенную по методу дробных шагов. Выражение (4.58) для модуля перехода показывает, что скорость затухания возмущений во всем спектре частот ω_1 и ω_2 может быть сделана сколь угодно большой при достаточно большом τ . Однако с увеличением τ возрастают и ошибки аппроксимации, связанные с представлением оператора перехода от слоя k к слою $k+1$ в виде произведения операторов, соответствующих «полушагам». В предельном случае $\tau = \infty$ получаем два слоя («целый» и «получелый»), не имеющие ничего общего с искомым решением и совсем непохожие друг на друга. Возникает естественная идея варьирования τ : сначала, когда преобладают возмущения, связанные с ошибками начального слоя, «гасить» побыстрее эти возмущения, а затем, когда начинают все большую роль играть ошибки аппроксимации, постепенно уменьшать τ . На основе этой идеи построены весьма эффективные алгоритмы для решения стационарных сеточных краевых задач.

В заключение сделаем одно замечание. При использовании метода установления иногда вычисления прекращают, когда удовлетворяется неравенство $|u^{(k+1)} - u^{(k)}| \leq \varepsilon$, где ε – заданная точность. Следует иметь в виду, что такой критерий недостаточно надежен. При медленном процессе установления это неравенство может быть выполнено, тогда как $|u^{(l+1)} - u|$ может быть еще достаточно велика. На практике можно воспользоваться следующим критерием окончания вычислений: $|u^{(l+1)} - u^{(k)}| < \varepsilon$, где $l-k$ достаточно велико (несколько десятков). Можно улучшить критерий окончания счета на установление, сделав его по трем последним итерациям (см. метод Эйткена п. 1.14).

4.16. МЕТОДЫ СКВОЗНОГО СЧЕТА

В задачах механики сплошных сред приходится иметь дело с течениями, в которых возникают поверхности сильного разрыва, такие как ударные волны, центрированные волны разрежения, поверхности контактного разрыва. На этих поверхностях терпят разрыв функции и их производные. При расчете таких те-

чений приходится либо явно выделять эти поверхности, используя законы сохранения, либо строить специальные методы решения, называемые методами сквозного счета, в которых априори неизвестные поверхности разрыва явно не выделяются и просчитываются «насквозь» вместе с областями непрерывного течения.

Среди методов сквозного счета можно выделить дивергентные схемы – схемы, обладающие аппроксимационной вязкостью. Одним весьма эффективным и красивым методом сквозного счета, использующим схемы с аппроксимационной вязкостью, является метод Годунова.

В основе метода Годунова лежат две идеи. Первая состоит в использовании при построении разностной схемы точных решений с кусочно-постоянными начальными данными. Для гиперболических уравнений такими точными решениями являются совокупности сравнительно простых и независимых решений задачи о распаде произвольного разрыва.

Вторая идея состоит в использовании гибких и деформирующихся разностных сеток, связанных с поверхностями разрывов и выделяемых при расчете начальной стадии. Рассмотрим применение метода Годунова для решения задачи Коши для уравнения акустики:

$$\begin{aligned}\frac{\partial U}{\partial t} + \frac{1}{\rho_0} \frac{\partial P}{\partial x} &= 0, \\ \frac{\partial P}{\partial t} + \rho_0 a_0^2 \frac{\partial U}{\partial x} &= 0,\end{aligned}\quad (4.61)$$

где U и P – соответственно возмущения скорости и давления в акустической волне, a_0 и ρ_0 – скорость звука и плотность в невозмущенном газе. Используя инвариант Римана $I_{\pm} = P \pm a_0 \rho_0 U$, для этих уравнений можно записать общее решение уравнения (4.61) в виде

$$\begin{aligned}U &= \frac{1}{2} [f(x - a_0 t) + g(x + a_0 t)], \\ P &= \frac{\rho_0 a_0}{2} [f(x - a_0 t) - g(x + a_0 t)],\end{aligned}\quad (4.62)$$

где f и g – произвольные функции, определяемые из начальных условий.

Построим с использованием (4.62) решение простейшей задачи о распаде. Пусть в момент $t = 0$ заданы следующие начальные условия:

$$\begin{aligned}U_0(x) &= U_I, P_0(x) = P_I \text{ при } x < x^*, \\ U_0(x) &= U_{II}, P_0(x) = P_{II} \text{ при } x > x^*,\end{aligned}\quad (4.63)$$

где U_I, U_{II}, P_I, P_{II} – некоторые константы, такие, что $U_I \neq U_{II}$ и $P_I \neq P_{II}$.

Начальные условия (4.63) таковы, что имеет место произвольный разрыв функций U и P в точке x^* , который с течением времени каким-то образом трансформируется в пространстве. Такая задача и называется задачей о распаде произвольного разрыва. Построим её решение, используя свойства инвариантов

Римана. Поскольку инвариант Римана I_+ постоянен вдоль характеристики $x - a_0 t = \text{const}$, а инвариант I_- постоянен вдоль характеристики $x + a_0 t = \text{const}$, приходим к структуре решения, изображенной на рис. 4.19. Нетрудно видеть, что решение задачи о распаде разрыва имеет вид

$$\begin{aligned}U &= U_I, P = P_I \text{ при } x < x^* - a_0 t \text{ в области I,} \\ U &= U_{II}, P = P_{II} \text{ при } x > x^* + a_0 t \text{ в области II,} \\ U &= \frac{U_I + U_{II}}{2} - \frac{P_{II} - P_I}{2a_0 \rho_0}, P = \frac{P_I + P_{II}}{2} - a_0 \rho_0 \frac{U_{II} - U_I}{2} \\ &\text{при } x^* - a_0 t < x < x^* + a_0 t \text{ в области III.}\end{aligned}\quad (4.64)$$

Эти области изображены на рис. 4.19.

Построенные функции $U(x, t)$ и $P(x, t)$ имеют разрыв вдоль характеристик $x + a_0 t = x^*$ и $x - a_0 t = x^*$, разделяющих области I, II и III. Поскольку функции U и P не являются непрерывными, то примем, что полученное решение является обобщенным решением задачи о распаде разрыва.

Перейдем к построению разностной схемы метода Годунова. Пусть в момент времени $t = 0$ заданы непрерывные функции $U_0(x)$ и $P_0(x)$. Заменим область непрерывного изменения аргумента дискретным множеством точек – разностной сеткой, узлы которой обозначим через x_m . Расстояние между соседними узлами есть шаг разностной сетки. Примем, что между точками x_{m-1} и x_m функции U и P постоянны. Такое предположение эквивалентно тому, что непрерывные функции $U_0(x)$ и $P_0(x)$ заменены некоторыми кусочно-постоянными функциями, сохраняющими постоянные значения между узлами разностной сетки (рис. 4.20, а).

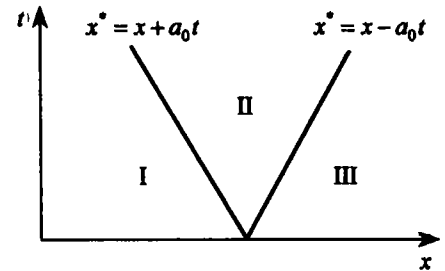


Рис. 4.19

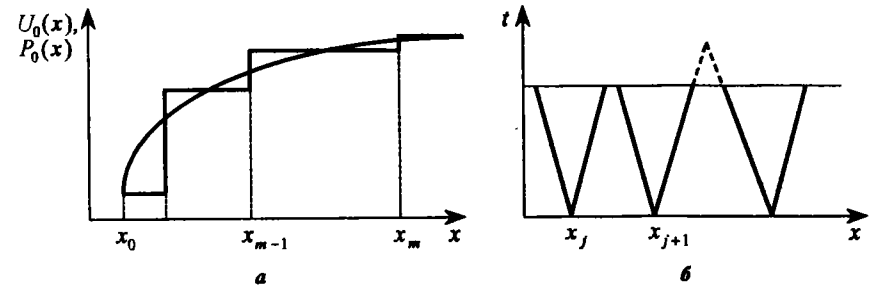


Рис. 4.20

Их значения между узлами разностной сетки обозначим $U_{m-1/2}$ и $P_{m-1/2}$, присвоив соответствующему слою полуцелый индекс $(m-1/2)$. В силу сказанного на границе между слоями возникает распад разрыва. В результате в каждом узле сетки образуются звуковые волны, распространяющиеся вправо и влево со скоростью звука a_0 . Через некоторое время τ структура решения имеет вид, изображенный на рис. 4.20, б. Согласно формулам (4.64), решение в окрестности точки x_m будет иметь вид

$$\begin{aligned} U &= U_{m-1/2}, \quad P = P_{m-1/2} \quad \text{в области I: } x_{m-1} + a_0 t < x < x_m - a_0 t, \\ U &= U_{m+1/2}, \quad P = P_{m+1/2} \quad \text{в области II: } x_m + a_0 t < x < x_{m+1} - a_0 t, \\ U &= U_m = \frac{U_{m-1/2} + U_{m+1/2}}{2} - \frac{P_{m+1/2} - P_{m-1/2}}{2a_0 \rho_0} \quad (4.65) \\ &\quad \text{в области III: } x_m - a_0 t < x < x_m + a_0 t, \\ P &= P_m = \frac{P_{m-1/2} + P_{m+1/2}}{2} - a_0 \rho_0 \frac{U_{m+1/2} - U_{m-1/2}}{2}. \end{aligned}$$

Здесь через U_m и P_m обозначены значения, возникающие при распаде разрыва в узле m .

Изображенная на рис. 4.20 структура решения сохранится до тех пор, пока звуковые волны, вышедшие из соседних узлов, не встретятся между собой. После этого её нужно перестроить заново, и даже в случае равномерной сетки эта процедура оказывается чрезвычайно громоздкой.

В связи с этим предлагается следующий способ построения решения в момент времени $t = \tau$. Заменяем решение (4.65) другим решением, таким, которое имеет ту же структуру, что и в начальный момент $t = 0$. Приближенно прием, что решение при $t = \tau$ также кусочно-постоянно внутри интервалов, ограниченных прежними узлами сетки x_m . Новые средние значения этих функций при $t = \tau$ между узлами x_{m-1} и x_m обозначим через $U^{m-1/2}$ и $P^{m-1/2}$.

Оказывается, что эти величины можно определить по формулам

$$\begin{aligned} U^{m-1/2} &= U_{m-1/2} - \frac{\tau}{h} (P_m - P_{m-1}), \\ P^{m-1/2} &= P_{m-1/2} - \frac{\tau}{h} \rho_0 a_0^2 (U_m - U_{m-1}), \end{aligned} \quad (4.66)$$

где U и P в правой части определяются по формулам (4.65) соответственно для точек m и $m-1$.

Действительно, рассмотрим законы сохранения в интегральной форме для первого и второго уравнений (4.61), которые следуют из формулы Грина

$$\iint \left(\frac{\partial \Phi_1}{\partial t} + \frac{\partial \Phi_2}{\partial x} \right) dx dt = \oint (\Phi_1 dx - \Phi_2 dt),$$

где $\Phi_1 = U$ или P и $\Phi_2 = P/\rho_0$ или $\rho_0 a_0^2 U$ для первого и второго уравнения соответственно. Имеем

$$\begin{aligned} \oint (U dx - (P/\rho_0) dt) &= 0, \\ \oint (P dx - \rho_0 a_0^2 U dt) &= 0. \end{aligned}$$

Запишем эти соотношения для прямоугольной ячейки в плоскости (x, t) , ограниченной прямыми $x = x_{m-1}$, $x = x_m$, $t = 0$, $t = \tau$. Имеем

$$\begin{aligned} \int_{x_{m-1}}^{x_m} U(x, t) dx &= \int_{x_{m-1}}^{x_m} U(x, 0) dx - \frac{1}{\rho_0} \int_0^\tau [P(x_m, t) - P(x_{m-1}, t)] dt, \\ \int_{x_{m-1}}^{x_m} P(x, t) dx &= \int_{x_{m-1}}^{x_m} P(x, 0) dx - \rho_0 a_0^2 \int_0^\tau [U(x_m, t) - U(x_{m-1}, t)] dt. \end{aligned} \quad (4.67)$$

Так как $U(x, 0)$ и $P(x, 0)$ постоянны на интервале от x_{m-1} до x_m и равны соответственно $U_{m-1/2}$ и $P_{m-1/2}$ и, аналогично, $U(x, \tau)$ и $P(x, \tau)$ равны $U^{m-1/2}$ и $P^{m-1/2}$, а на левой и правой сторонах ячейки $U(x, t)$ и $P(x, t)$ принимают постоянные значения U_{m-1} , U_m и P_{m-1} , P_m соответственно, которые вырабатываются при распаде разрыва в узлах m и $(m-1)$, то из (4.67) имеем (4.66).

Очевидно, что формулы (4.67) справедливы до тех пор, пока не происходит пересечения звуковых волн, приходящих из соседних ячеек. Это имеет место при $\tau < h/a_0$, что, как видно, есть условие КФЛ. Описанная разностная схема перехода с одного временного слоя на другой имеет первый порядок точности по τ и h .

4.17. МЕТОД ПРЯМЫХ

Основная идея метода прямых и рассматриваемого далее метода характеристик состоит в сведении уравнений в частных производных к решению системы обыкновенных дифференциальных уравнений. В этом состоит их отличие от метода сеток, который непосредственно сводит решение систем уравнений в частных производных к решению систем алгебраических уравнений. Метод прямых может быть использован для решения уравнений в частных производных любого типа, но используется в основном для решения эллиптических и параболических уравнений.

В методе прямых уравнения в частных производных рассматриваются на линиях, параллельных одной из координатных осей, что приводит к исчезновению зависимости от одной из независимых переменных. Сущность метода состоит в следующем. Пусть в прямоугольной области $G = [\alpha < x < \beta, y_0 < y < y_0 + H]$ необходимо найти решение дифференциального уравнения

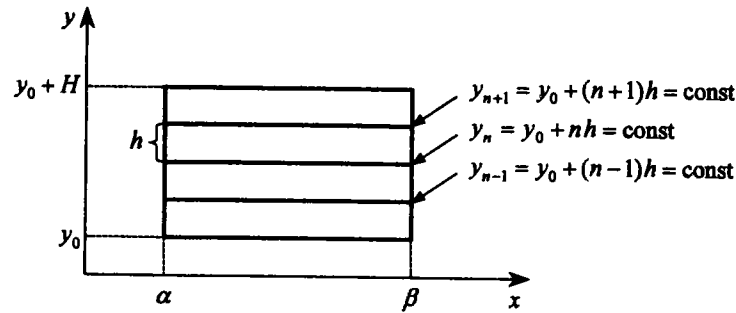


Рис 4.21

$$a(x, y) \frac{\partial^2 u}{\partial x^2} + b(x, y) \frac{\partial^2 u}{\partial y^2} + c(x, y) \frac{\partial u}{\partial x} + d(x, y) \frac{\partial u}{\partial y} + e(x, y) u = f(x, y), \quad (4.68)$$

где $a, b > 0$. Пусть уравнение (4.68) удовлетворяет следующим граничным условиям:

$$\begin{aligned} u(x, y_0) = \varphi_0(x), \quad u(x, y_0 + H) = \varphi_1(x) \quad \text{при } \alpha \leq x \leq \beta, \\ u(\alpha, y) = \psi_0(y), \quad u(\beta, y) = \psi_1(y) \quad \text{при } y_0 < y < y_0 + H, \end{aligned} \quad (4.69)$$

где $\varphi_j(x)$, $\psi_j(y)$, $j = 0, 1$ – заданные функции (рис. 4.21).

На отрезке $[y_0, y_0 + H]$ возьмем точки $y_n = y_0 + nh$, $n = \overline{1, N}$, $h = H/(1 + N)$ и проведем прямые $y = y_n = \text{const}$ (рис. 4.21). Предполагая существование достаточно гладкого решения $u(x, y)$ задачи (4.68) – (4.69), положим в уравнении (4.68) $y = y_n$ и заменим производные по y разностными отношениями, воспользовавшись, например, соотношениями

$$\begin{aligned} \left. \frac{\partial u}{\partial y} \right|_{y=y_n} &= \frac{1}{2h} [u_{n+1}(x) - u_{n-1}(x)] + o(h^2), \\ \left. \frac{\partial^2 u}{\partial y^2} \right|_{y=y_n} &= \frac{1}{h^2} [u_{n+1}(x) - 2u_n(x) + u_{n-1}(x)] + o(h^2), \quad \text{где } u_n(x) = u(x, y_n). \end{aligned}$$

Проводя такую замену на всех линиях $y_n = \text{const}$ ($n = \overline{1, N}$) и пренебрегая малыми порядка $O(h^2)$, получаем систему N обыкновенных линейных дифференциальных уравнений второго порядка:

$$\begin{aligned} a_n(x) u_n''(x) + \frac{b_n(x)}{h^2} [u_{n+1}(x) - u_n(x) + u_{n-1}(x)] + c_n(x) u_n'(x) + \frac{d_n(x)}{2h} \times \\ \times [u_{n+1}(x) - u_{n-1}(x)] + e_n(x) u_n(x) = f_n(x), \quad n = \overline{1, N}, \end{aligned} \quad (4.70)$$

где $u_n(x)$ обозначает приближенное значение решения $u(x, y)$ на прямой $y = y_n$; а $a_n(x)$, $b_n(x)$, $c_n(x)$, $d_n(x)$, $e_n(x)$ и $f_n(x)$ – значения коэффициентов a , b , c , d , e , f в уравнении (4.68) на линии $y = y_n = \text{const}$.

Используя граничные условия на границе области, имеем

$$\begin{aligned} u_0(x) &= \varphi_0(x), \quad \alpha \leq x \leq \beta, \\ &\dots \\ u_{n+1}(x) &= \varphi_{n+2}(x), \quad \alpha \leq x \leq \beta, \\ u_n(\alpha) &= \psi_0(y_n), \quad u_n(\beta) = \psi_1(y_n), \quad n = \overline{1, N}. \end{aligned} \quad (4.71)$$

На прямой $y = y_n$ каждое из уравнений (4.70) содержит три неизвестных значения функции $u_{n+1}(x)$, $u_n(x)$, $u_{n-1}(x)$. Всего с учетом граничных значений число неизвестных равно числу линий $y_n = \text{const}$ и равно числу уравнений в системе

Система (4.70) обыкновенных дифференциальных уравнений с граничными условиями (4.71) аппроксимирует уравнение (4.68) с граничными условиями (4.69) и называется системой уравнений метода прямых.

Эта система уравнений вместе с граничными условиями решается численно (см. гл. 3, пп. 3.8 – 3.10), либо аналитически.

В зависимости от способа замены производных по y разностными соотношениями будем иметь различные системы метода прямых, с различной точностью аппроксимирующие дифференциальное уравнение (4.68). Метод прямых удобнее всего применять в том случае, когда коэффициенты в уравнении (4.68) не зависят от x . В этом случае система (4.70) будет системой обыкновенных дифференциальных уравнений с постоянными коэффициентами.

Метод прямых можно рассматривать как предельный случай метода сеток, если шаг сетки по оси x устремить к нулю. Отметим важное свойство метода прямых. Если изменение искомой функции внутри области незначительно, то можно ограничиться разбиением области на одну или две полосы, что в ряде случаев существенно упрощает задачу, а иногда позволяет получить аналитическое решение. В механике сплошных сред широко применяется метод интегральных соотношений, аналогичный методу прямых.

4.18. МЕТОД ХАРАКТЕРИСТИК

Метод характеристик применяется для решения уравнений гиперболического типа в задачах газовой динамики и теории упругости. Метод характеристик позволяет уменьшать на единицу размерность задачи и в случае уравнений в частных производных с двумя независимыми переменными сводит их к системе обыкновенных дифференциальных уравнений. Метод характеристик тесно связан с решением задачи Коши. Характеристические поверхности (линии) определяются как поверхности (линии), на которых решение задачи Коши либо не существует, либо не единственно.

Рассмотрим квазилинейную систему порядка P вида

$$A \frac{\partial u}{\partial x} + B \frac{\partial u}{\partial y} = f, \quad (4.72)$$

где матрицы A и B и вектор f с элементами a_{ij}, b_{ij}, f_i есть функции независимых переменных x, y и компонент $u_1, \dots, u_P$ искомой вектор-функции. Оказывается, что в плоскости (x, y) существуют линии, называемые характеристиками, на которых решение задачи либо не существует, либо не единственно. Уравнение этих линий можно получить, приравняв нулю определитель

$$\left| B - A \frac{\partial y}{\partial x} \right| = 0 \quad (4.73)$$

Из соотношения (4.73), которое является многочленом степени P , можно найти производные dy/dx , определяющие характеристические направления. Таким образом, получаем P дифференциальных уравнений

$$\frac{dy}{dx} = \lambda_i(x, y, u), \quad i = \overline{1, P}, \quad (4.74)$$

которые называют уравнениями направлений или уравнениями характеристик. Отметим, что p действительных корней уравнения (4.73) существует только для гиперболических систем.

Через каждую точку плоскости (x, y) проходит p характеристик. На характеристиках выполняются соотношения, называемые уравнениями совместности. Эти соотношения связывают дифференциалы искомых функций с дифференциалами независимых переменных вдоль характеристик и получаются, если приравнять нулю определитель порядка p расширенной матрицы $(Bdx - Adu, fdx - Adu)$.

Таким образом, можно видеть, что система дифференциальных уравнений в частных производных (4.72) сводится к эквивалентной системе обыкновенных дифференциальных уравнений.

Поясним сказанное на примере системы уравнений в частных производных с двумя неизвестными функциями u_1 и u_2 . Имеем систему уравнений

$$A \frac{\partial u}{\partial x} + B \frac{\partial u}{\partial y} = f, \quad (4.75)$$

где A и B – матрицы второго порядка, $u = (u_1, u_2)^T$ – вектор-столбец, $f = (f_1, f_2)^T$ – вектор-столбец правых частей. Перепишем систему (4.75) в виде

$$\begin{aligned} a_{11} \frac{\partial u_1}{\partial x} + a_{12} \frac{\partial u_2}{\partial x} + b_{11} \frac{\partial u_1}{\partial y} + b_{12} \frac{\partial u_2}{\partial y} &= f_1, \\ a_{21} \frac{\partial u_1}{\partial x} + a_{22} \frac{\partial u_2}{\partial x} + b_{21} \frac{\partial u_1}{\partial y} + b_{22} \frac{\partial u_2}{\partial y} &= f_2 \end{aligned} \quad (4.76)$$

Уравнение (4.73) для системы (4.76) имеет вид

$$\begin{vmatrix} b_{11} - \lambda a_{11} & b_{12} - \lambda a_{12} \\ b_{21} - \lambda a_{21} & b_{22} - \lambda a_{22} \end{vmatrix} = 0 \quad \text{или} \quad a\lambda^2 + c\lambda + b = 0,$$

$$\text{где } a = |A|, \quad b = |B|, \quad c = \begin{vmatrix} b_{12} & b_{21} \\ a_{12} & a_{22} \end{vmatrix} - \begin{vmatrix} a_{11} & b_{11} \\ a_{12} & b_{22} \end{vmatrix}.$$

Решив это уравнение, найдем два уравнения для направлений двух характеристик, проходящих через каждую точку плоскости (x, y) (рис. 4.22, а):

$$\frac{dy}{dx} = \lambda^{(1)}(x, y, u_1, u_2) = \frac{-c + \sqrt{c^2 - 4ab}}{2a}, \quad (4.77)$$

$$\frac{dy}{dx} = \lambda^{(2)}(x, y, u_1, u_2) = \frac{-c - \sqrt{c^2 - 4ab}}{2a}. \quad (4.78)$$

Уравнения совместности получим, если приравняем нулю определители первого и третьего или второго и третьего столбцов расширенной матрицы

$$\begin{vmatrix} b_{11}dx - a_{11}dy & b_{12}dx - a_{12}dy & f_1dx - a_{11}du_1 - a_{12}du_2 \\ b_{21}dx - a_{21}dy & b_{22}dx - a_{22}dy & f_2dx - a_{21}du_1 - a_{22}du_2 \end{vmatrix}.$$

Уравнения совместности имеют вид

$$[a\lambda^{(k)} + b]du_1 + \bar{c}du_2 + hdx + e dy = 0, \quad k = 1, 2, \quad hdx + e dy = (h + e\lambda^{(k)})dx \quad (4.79)$$

где $\bar{c} = a_{22}b_{12} - a_{12}b_{22}$, $h = b_{22}f_1 - b_{12}f_2$, $e = a_{12}f_2 - a_{22}f_1$, $g = h + e\lambda^{(k)}$.

Уравнения (4.77) и (4.79) образуют систему четырех обыкновенных дифференциальных уравнений для нахождения значений x, y, u_1, u_2 вдоль характеристик.

Перейдем к изложению существа численного метода характеристик. Рассмотрим в плоскости (x, y) две достаточно близкие точки 1 и 2 (рис. 4.22, б), не лежащие на одной характеристике, и зададимся целью определить значения u_1 и u_2 в точке 3 пересечения характеристик, проходящих через эти точки. Координаты точки 1 будем обозначать (x_1, y_1) , а значение любой функции q в точке 1 – через q_1 .

Переходя в (4.77) и (4.78) от дифференциалов к разностям, имеем для определения координат точки 3 (x_3, y_3) систему двух линейных уравнений:

$$\begin{cases} y_3 - y_1 = \lambda_1^{(1)}(x_3 - x_1), \\ y_3 - y_2 = \lambda_2^{(2)}(x_3 - x_2). \end{cases}$$

Аналогично из (4.79) получаем

$$(\lambda^{(1)}a + b)_1(u_{13} - u_{11}) + c_1(u_{23} - u_{21}) + g_1(x_3 - x_1) = 0,$$

$$(\lambda^{(2)}a + b)_2(u_{13} - u_{12}) + c_2(u_{23} - u_{22}) + g_2(x_3 - x_2) = 0,$$

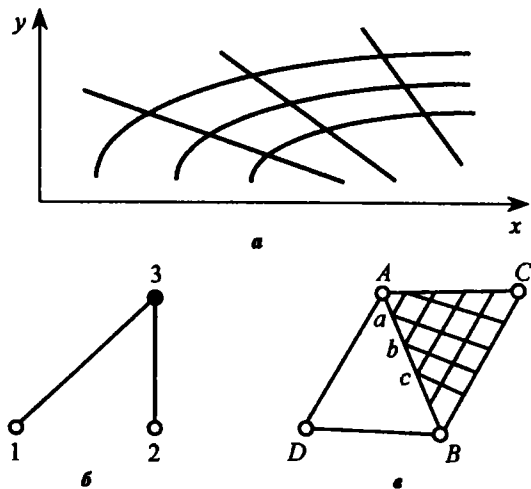


Рис. 4.22

где $u_j = u_j(x_j, y_j)$. Решая эту систему относительно x_3, y_3, u_{13} и u_{23} , находим приближенные значения x_3, y_3 и u_1, u_2 в точке 3. Это приближение может оказаться недостаточно точным, так как мы заменили характеристики, выходящие из точек 1 и 2, отрезками прямых. В действительности точка 3 должна быть точкой пересечения, вообще говоря, криволинейных характеристик, и, кроме того, дифференциалы всюду мы заменили разностями. Ошибка в определении x_3, y_3, u_{13} и u_{23} будет порядка квадрата расстояния от точки 3 до точек 1 и 2. Решение можно уточнить, если при вычислении коэффициентов при дифференциалах в (4.77) – (4.79) использовать значения искомых величин в точке 3. Этого можно достичь, используя следующий итерационный процесс:

$$\begin{aligned} y_3^{(k)} - y_1 &= (\lambda^{(1)})_1^{(k)} (x_3^{(k)} - x_1), \\ y_3^{(k)} - y_2 &= (\lambda^{(2)})_2^{(k)} (x_3^{(k)} - x_2), \\ (\lambda^{(1)}a + b)_1^{(k)} (u_{1,3}^{(k)} - u_{1,1}) + c_1^{(k)} (u_{2,3}^{(k)} - u_{2,1}) + g_1^{(k)} (x_3^{(k)} - x_1) &= 0, \\ (\lambda^{(2)}a + b)_2^{(k)} (u_{1,3}^{(k)} - u_{1,2}) + c_2^{(k)} (u_{2,3}^{(k)} - u_{2,2}) + g_2^{(k)} (x_3^{(k)} - x_2) &= 0, \end{aligned}$$

где $k=1,2,3,\dots$ – номер итерации, $q_i^{(k)} = (q_i^{(k-1)} + q_i)/2$ – общая формула для определения значений коэффициентов уравнений при переходе с итерации на итерацию, при этом первая итерация соответствует методу Эйлера, а вторая – методу Эйлера с пересчетом.

Пусть теперь u_1 и u_2 заданы на дуге AB некоторой кривой, которая ни в одной точке не имеет характеристического направления. Требуется определить решение в окрестности AB . Эта задача – задача Коши. Выберем на AB ряд точек

$A, a, b, \dots, B$ (рис. 4.22, в) и проведем через каждую из них характеристики обоих семейств (4.77) и (4.78). В точках их пересечения описанным выше способом можно вычислить искомые функции. Зная решение в этих точках, можно вычислить решение еще на одном слое, и так до тех пор, пока не найдем решение в точке C . Таким образом последовательно определяется решение и одновременно выстраивается характеристическая сетка. Аналогично определяется решение в характеристическом треугольнике ABD (рис. 4.22, в).

4.19. МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ

В последние годы для решения краевых задач механики сплошной среды интенсивно развивается метод конечных элементов. Этот метод начал использоваться для задач строительной механики, а затем превратился в общий численный метод решения систем обыкновенных дифференциальных уравнений и уравнений в частных производных.

Метод конечных элементов содержит в себе основные концепции метода сеток, связанные с дискретизацией областей непрерывного изменения аргумента и искомой функции, и метода Галеркина, использующего вариационные принципы для нахождения искомых функций в узлах рассматриваемой области. В конечном итоге решение краевой задачи сводится к решению системы линейных алгебраических уравнений, обладающих ленточной матрицей. Порядок системы равен числу узловых точек.

Основная идея метода конечных элементов состоит в том, что любую непрерывную функцию, такую как зависимость температуры, давления, перемещения от координат, можно аппроксимировать дискретной моделью, которая строится на множестве кусочно-непрерывных функций, определенных на конечном числе подобластей. Кусочно-непрерывные функции определяются с помощью значений непрерывной функции в конечном числе точек рассматриваемой области.

В общем случае значения непрерывной функции заранее неизвестны и нужно определить их в некоторых внутренних точках области. Дискретную модель, однако, легко построить, если сначала предположить, что числовые значения искомой функции в каждой внутренней точке области известны.

Построение дискретной модели состоит из следующих этапов:

1. В рассматриваемой области фиксируется конечное число точек. Эти точки называются узловыми точками или просто узлами.
2. Значения непрерывной функции в каждой узловой точке считаются неизвестными, которые должны быть определены.
3. Область определения разбивается на конечное число подобластей, называемых элементами. Эти элементы имеют общие узловые точки и в совокупности аппроксимируют форму области.
4. Непрерывная функция аппроксимируется на каждом элементе полиномом, который определяется с помощью узловых значений функции. Для каждо-

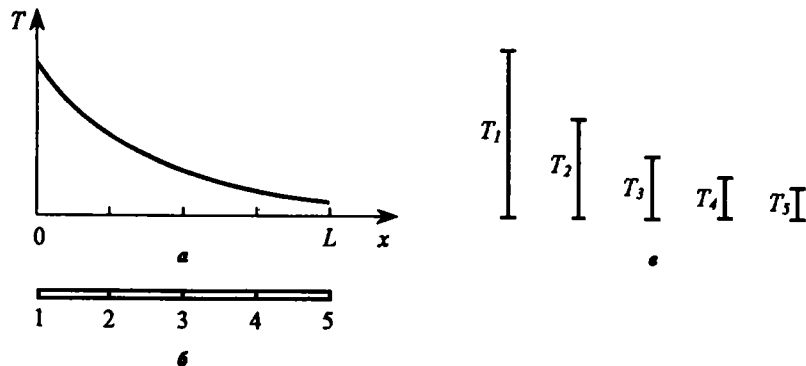


Рис. 4.23

го элемента определяется свой полином, но полиномы подбираются таким образом, чтобы сохранялась непрерывность функции вдоль границы.

Основные концепции метода конечных элементов могут быть наглядно проиллюстрированы на примере заданного одномерного распределения температуры в стержне (рис. 4.23, а). Рассматриваемая непрерывная функция – $T(x)$, область определения – отрезок OL вдоль оси x . Фиксированы и пронумерованы пять точек на оси x (рис. 4.23, б). Эти узловые точки совсем не обязательно располагать на равном расстоянии друг от друга. Очевидно, можно ввести в рассмотрение и более пяти точек, но этих пяти вполне достаточно, чтобы проиллюстрировать основную идею метода. Значения T в данном случае известны в каждой узловой точке. Эти фиксированные значения представлены графически на рис. 4.23, в и обозначены в соответствии с номерами узловых точек.

Разбиение области на элементы может быть проведено различными способами. Можно, например, ограничить каждый элемент двумя соседними узловыми точками, образовав четыре элемента (рис. 4.24, а), или разбить на два элемента, каждый из которых содержит три узла (рис. 4.24, б).

Соответствующий элементу полином определяется по значению функции T в узловых точках элемента. В случае разбиения области на четыре элемента, когда на каждый элемент приходится по два узла, функция элемента будет линейна по T (две точки однозначно определяют прямую линию): окончательная аппроксимация T будет состоять из четырех кусочно-линейных функций, каждая из которых определена на отдельном элементе. В случае разбиения области на два элемента с тремя узловыми точками каждый функция элемента представляет собой полином второй степени: окончательная аппроксимация T будет состоять из двух кусочно-непрерывных квадратичных функций. Отметим, что и в том и в другом случае приближение будет именно кусочно-непрерывным, так как углы наклона аппроксимирующих полиномов в граничных угловых точках, принадлежащих двум элементам, могут иметь различные значения.

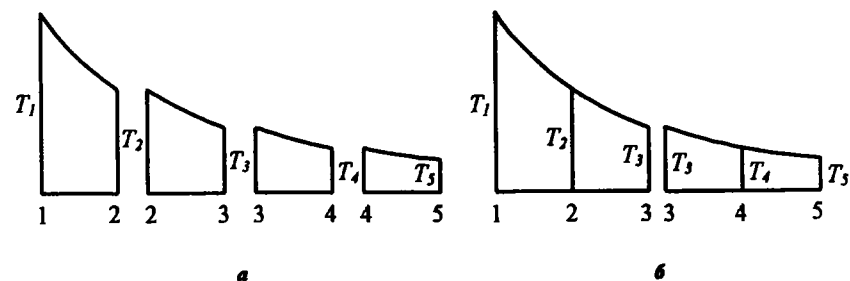


Рис. 4.24

В общем случае распределение температуры неизвестно и мы хотим определить её значение в некоторых фиксированных точках. Методика построения дискретной модели остается точно такой же, как описано выше, но с добавлением одного дополнительного шага. Снова определяется множество узлов и значения температуры в этих узлах, но теперь эти значения являются неизвестными. Область решения разбивается на элементы, на каждом из которых определяется соответствующая функция элемента. Узловые значения T должны быть теперь определены таким образом, чтобы обеспечивалось в каком-то смысле наилучшее приближение к истинному распределению температуры. Это наилучшее приближение находится из условия минимизации некоторой величины, связанной с физической сущностью задачи. Если рассматривается задача о распределении тепла, то минимизируется функционал, связанный с соответствующим дифференциальным уравнением. Процесс минимизации сводится к решению системы линейных алгебраических уравнений относительно узловых значений T .

При построении дискретной модели непрерывной величины, определенной в двух- или трехмерной области, основная концепция метода конечных элементов используется аналогично. В двумерном случае элементы описываются функциями от x и y . При этом чаще всего рассматриваются элементы в форме треугольника или четырехугольника. Функции элемента представляются теперь плоскими или криволинейными поверхностями: функция элемента представляется плоскостью, если для данного элемента взято минимальное число точек, которое для треугольного элемента равно трем, а для четырехугольного – четырем.

Важным аспектом метода конечных элементов является возможность выделить из набора элементов типичный элемент при нахождении функции элемента. Это позволяет определить функцию элемента независимо от относительного положения элемента в общей связанной модели и от других функций элементов. Задание функции элемента через произвольное множество узловых значений и координат дает возможность использовать функции элемента для аппроксимации искомой функции на выбранных узлах.

Простейшим из элементов является одномерный элемент, схематически изображенный в виде отрезка. Простейший одномерный элемент имеет два узла, по одному на каждом конце элемента. Одномерные элементы более высокого порядка, трехузловые (квадратичные) и четырехузловые (кубические), содержат по три и четыре узла соответственно. Порядок элемента, как будет показано ниже, определяется порядком интерполяционного многочлена, с помощью которого аппроксимируется искомая функция. Так на рис. 4.24, а представлены одномерные линейные элементы, а на рис. 4.24, б – квадратичные. В качестве двумерных элементов используются треугольники и четырехугольники. При этом количество узлов, которые содержит элемент, определяет его порядок и порядок его интерполяционного многочлена.

В соответствии с вышесказанным искомую функцию φ записывают на элементе в виде интерполяционного многочлена. В случае линейного одномерного элемента, содержащего два узла (так называемый одномерный симплекс-элемент), интерполяционный многочлен имеет вид

$$\varphi = \alpha_1 + \alpha_2 x,$$

а в случае двумерного линейного элемента, содержащего три узла (так называемый двумерный симплекс-элемент) интерполяционный многочлен имеет вид

$$\varphi = \alpha_1 + \alpha_2 x + \alpha_3 y.$$

В случае квадратичного одномерного элемента (так называемый одномерный комплекс-элемент) интерполяционный многочлен имеет вид

$$\varphi = \alpha_1 + \alpha_2 x + \alpha_3 x^2,$$

а в случае квадратичного двумерного элемента (так называемый двумерный комплекс-элемент) –

$$\varphi = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2.$$

Входящие в интерполяционные формулы коэффициенты α_i выражаются через значения функции в узлах. В частности, для одномерного симплекс-элемента

$$\varphi = N_i \Phi_i + N_j \Phi_j = [N] \{\Phi\},$$

где N_i и N_j – так называемые функции формы, определяемые соотношениями

$$N_i = \frac{(X_j - x)}{L}, \quad N_j = \frac{(x - X_i)}{L},$$

где X_i и X_j – координаты узлов элемента, L – длина элемента, Φ_i и Φ_j – значения искомых функций в узлах i и j , $[N] = [N_i, N_j]$ – матричная строка, а $\{\Phi\} = \{\Phi_i, \Phi_j\}$ – вектор-столбец искомых функций.

Для двумерного комплекс-элемента

$$\varphi = N_i \Phi_i + N_j \Phi_j + N_k \Phi_k = [N] \{\Phi\}, \quad (4.80)$$

где

$$\begin{aligned} N_i &= \frac{1}{2A}(a_i + b_i x + c_i y), & a_i &= X_j Y_k - X_k Y_j, \\ N_j &= \frac{1}{2A}(a_j + b_j x + c_j y), & b_i &= Y_j - Y_k, \\ N_k &= \frac{1}{2A}(a_k + b_k x + c_k y), & c_i &= X_k - X_j, \\ A &= \begin{vmatrix} 1 & X_i & Y_i \\ 1 & X_j & Y_j \\ 1 & X_k & Y_k \end{vmatrix}, & a_j &= X_k Y_i - X_i Y_k, \\ & & b_j &= Y_k - Y_i, \\ & & c_j &= X_i - X_k, \\ [N] &= [N_i, N_j, N_k], & a_k &= X_i Y_j - X_j Y_i, \\ \{\Phi\} &= \{\Phi_i, \Phi_j, \Phi_k\}, & b_k &= Y_i - Y_j, \\ & & c_k &= X_j - X_i. \end{aligned}$$

Отметим, что функция формы N_p , равная единице в узле, равна нулю во всех остальных узлах в соответствии с построением интерполяционных многочленов.

Следующим этапом метода конечных элементов является получение системы уравнений для нахождения неизвестных значений функций в узлах. С этой целью данному дифференциальному уравнению с заданными граничными условиями ставится в соответствие некоторый функционал, минимум которого достигается в том случае, когда удовлетворяется исходное дифференциальное уравнение. Иными словами, вариационным уравнением Эйлера для данного функционала является наше исходное уравнение. Например, задача Дирихле для уравнения Лапласа с целью отыскания потенциала скорости

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0$$

эквивалентна отысканию минимума функционала

$$\chi = \frac{1}{2} \int_G \left[\left(\frac{\partial \varphi}{\partial x} \right)^2 + \left(\frac{\partial \varphi}{\partial y} \right)^2 \right] ds.$$

Минимизация этого функционала должна быть осуществлена на множестве узловых значений $\{\Phi\}$. Перед процедурой минимизации проведем некоторые преобразования. Представим функционал в виде

$$\chi = \frac{1}{2} \int_G \{g\}^T \{g\} ds, \quad (4.81)$$

где $\{g\}^T = \left\{ \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y} \right\}$. Рассмотрим область G , в которой ищется решение на E элементах и заменим интеграл (4.81) суммой интегралов по этим элементам. Имеем

$$\chi = \sum_{s=1}^E \frac{1}{2} \int_G \{g^s\}^T \{g^s\} ds = \chi^{(1)} + \chi^{(2)} + \dots + \chi^{(e)}.$$

Потребуем теперь, чтобы функционал χ был минимальным на множестве узловых значений $\{\Phi\}$. Для этого необходимо, чтобы

$$\frac{\partial \chi}{\partial \{\Phi\}} = \frac{\partial}{\partial \{\Phi\}} \sum_{s=1}^E \chi^{(s)} = \sum_{s=1}^E \frac{\partial \chi^{(s)}}{\partial \{\Phi\}} = 0.$$

Поскольку, согласно (4.80) $\varphi^{(s)} = [N^{(s)}]\{\Phi\}$, то

$$\{g^{(s)}\} = \begin{pmatrix} \frac{\partial \varphi^{(s)}}{\partial x} \\ \frac{\partial \varphi^{(s)}}{\partial y} \end{pmatrix} = \begin{bmatrix} \frac{\partial N_1^{(s)}}{\partial x} & \frac{\partial N_2^{(s)}}{\partial x} & \dots & \frac{\partial N_p^{(s)}}{\partial x} \\ \frac{\partial N_1^{(s)}}{\partial y} & \frac{\partial N_2^{(s)}}{\partial y} & \dots & \frac{\partial N_p^{(s)}}{\partial y} \end{bmatrix} \begin{Bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_p \end{Bmatrix}, \quad (4.82)$$

или

$$\{g^{(s)}\} = [B^{(s)}]\{\Phi\}, \quad (4.83)$$

где $B^{(s)}$ – матрица, содержащая информацию, связанную с функцией формы. Учитывая (4.82) и (4.83), функционал χ можно переписать в виде

$$\chi^{(s)} = \frac{1}{2} \int_G \{\Phi\}^T [B^{(s)}]^T [B^{(s)}] \{\Phi\} ds.$$

Окончательно записываем условие минимума функционала

$$\frac{\partial \chi}{\partial \{\Phi\}} = \sum_{s=1}^E [K^{(s)}]\{\Phi\} = 0, \quad (4.84)$$

где $K^{(s)} = \int_G [B^{(s)}]^T [B^{(s)}] ds$. Из (4.80) следует, что матрицы B и $K^{(s)}$ определяются следующими соотношениями для двумерного комплекс-элемента:

$$B = \frac{1}{2A} \begin{bmatrix} b_i & b_j & b_k \\ c_i & c_j & c_k \end{bmatrix},$$

$$K^{(s)} = \frac{1}{4A} \begin{bmatrix} b_i^2 & b_i b_j & b_i b_k \\ b_j b_i & b_j^2 & b_j b_k \\ b_k b_i & b_k b_j & b_k^2 \end{bmatrix} + \frac{1}{4A} \begin{bmatrix} c_i^2 & c_i c_j & c_i c_k \\ c_j c_i & c_j^2 & c_j c_k \\ c_k c_i & c_k c_j & c_k^2 \end{bmatrix}.$$

Уравнения (4.84) представляют собой систему линейных алгебраических уравнений относительно значений искомой функции в узлах $\{\Phi_1, \Phi_p, \dots, \Phi_p\}$. Таким образом, решение краевой задачи для уравнения в частных производных сведено к решению системы линейных алгебраических уравнений. Матрица этой системы является матрицей ленточного типа, содержащей большое число нулевых элементов. Матрица системы будет трехдиагональной, если используются двумерные симплекс-элементы. При использовании квадратичных элементов и элементов более высокого порядка матрица системы будет содержать большое число «диагоналей». Матрица системы является хорошо обусловленной.

Важным свойством матрицы является её симметрия и положительная определенность. Симметрия матрицы позволяет не хранить в памяти примерно половину её элементов. Положительная определенность матрицы означает, что элементы, стоящие на главной диагонали, всегда положительны и обычно много больше по величине, чем любой коэффициент соответствующей строки или столбца. Для симметричных положительно определенных матриц удается значительно сократить время вычислений и уменьшить влияние ошибок округления. Для решения систем линейных уравнений такого типа эффективным является метод Гаусса и его различные варианты, в том числе метод прогонки.

ЛИТЕРАТУРА

1. Бахвалов Н.С. Численные методы. – М.: Наука, 1973.
2. Волков Е.А. Численные методы. – М.: Наука, 1987.
3. Демидович Б.П., Марон И.А. Основы вычислительной математики. – М.: Наука, 1970.
4. Калиткин Н.Н. Численные методы. – М.: Наука, 1976.
5. Пирумов У.Г., Росляков Г.С. Численные методы газовой динамики. – М.: Высшая школа, 1987.
6. Роуч П. Вычислительная гидродинамика. – М.: Мир, 1980.
7. Самарский А.А., Попов Ю.П. Разностные схемы газовой динамики. – М.: Наука, 1975.
8. Самарский А.А. Введение в численные методы. – М.: Наука, 1982.
9. Самарский А.А. Теория разностных схем. – М.: Наука, 1977.
10. Самарский А.А., Николаев Е.С. Методы решения сеточных уравнений. – М.: Наука, 1978.
11. Сегерлинд Л. Применение метода конечных элементов. – М.: Мир, 1979.
12. Крысицкий В. Шеренга великих математиков. – Варшава, 1981.
13. Сборник задач по методам вычислений / Под ред. П.И.Монастырного. – М.: Изд-во «физ.-мат. лит-ра». ВО Наука, 1994.

ВЕЛИКИЕ МАТЕМАТИКИ



ЖАК САЛОМОН АДАМАР

(1865 – 1963)

Высшее образование Адамар получил в Париже. В 1892 году ему была присвоена степень доктора философии. В 1896 году он дал полное доказательство закона распределения простых чисел, на основе которого можно было с большой точностью определить количество простых чисел в любом промежутке. В 1893 – 1897 годах Адамар работал в университете в Бордо.

В 1897 году Адамар был приглашен на работу в Сорбонну, где оставался до 1901 года, после чего стал преподавать в College de France. В 1912 году он был избран членом Французской академии наук; в том же году, не бросая предыдущей работы, начал заниматься педагогической деятельностью и стал читать лекции в Ecole Polytechnique, вплоть до 1937 года.

Адамар был решительным противником всякого рода ограничений в выборе предмета исследований и методов математического действия. Он написал обширный и очень ценный учебник геометрии для средних школ. Занимался теорией чисел, дифференциальными уравнениями, теорией аналитических функций, уравнениями математической физики. Адамар был одним из основателей функционального анализа. В курсе математического анализа известна теорема о радиусе сходимости степенного ряда, носящая название теоремы Адамара-Коши, а в теории определителей – неравенства Адамара для абсолютной величины определителя. Умер Адамар в октябре 1963 года



ЮЗЕФ ГЁНЕ-ВРОНСКИЙ

(1778 – 1853)

В конце XIX века в историю науки вошли три польских ученых, внесшие значительный вклад в развитие мировой математики. Это – Адам Коханский, Ян Снядецкий и Юзеф Вронский.

Самым выдающимся из них был несомненно Вронский. Этот ученый родился в 1776 году в городе Вольштин в Великопольше. Его отец Гене, чех по происхождению, был образованным и богатым человеком.

Юзеф Гёне-Вронский окончил в Варшаве Кадетский корпус и получил звание поручика национальной артиллерии. Он отличился в боях против прусских войск, осаждавших Варшаву. Принимал участие в битве под Мацевевицами, во время которой попал в плен. В 1795 году поступил в царскую армию, служил в ней два года в звании майора, затем подполковника. Что склонило молодого Вронского решиться на такой шаг – неизвестно. Известно, однако, что, услышав об организации польских легионов, он вышел из русской армии и уехал в Германию. Находясь в Германии в течение двух лет, он близко познакомился с бурно развивавшейся в то время германской философией. В 1800 году Вронский уехал во Францию. После краткого пребывания в Париже он выехал в Марсель и там поступил в польский легион. Жизнь в Марселе оказала на судьбу Вронского большое влияние. Он окончательно оставил мысль о военной или дипломатической карьере и решил целиком посвятить себя научной работе.

Вронский занимался различными науками. В Марселе он написал свои первые математические труды, здесь же возникли первые наброски его будущих, написанных исключительно по-французски, работ, направленных на реформу человеческих знаний. В математике Вронский стремился связать воедино все её ветви, причем в его трудах можно усмотреть некоторые начала функционального анализа. Почести, которые оказала Вронскому марсельская общественность (он был избран членом-корреспондентом Научного общества в Марселе и генеральным секретарем Общества врачей) не удовлетворили его честолюбия, и он решил расширить поле своей деятельности. С этой целью во второй половине 1810 года он уехал в Париж, где и поселился. С этого времени он полностью посвятил себя работе. «Его железный организм, – пишет Дикштейн в биографии Вронского, насчитывающей 300 страниц, – требовал мало сна и питания; работу он начинал ранним утром и только после нескольких часов занятий принимался за скромный завтрак, говоря: «свой день я уже заработал».

Вронский оставил после себя большое количество работ по математике, философии, физике и техническим наукам. Группа его друзей пыталась после смерти ученого издать полное собрание его сочинений (многие из них остались

в рукописях), но оказалось, что такое издание состояло бы из десяти томов по восемьсот страниц каждый. Спустя тринадцать лет после смерти Вронского Польское общество точных наук в Париже, целью которого было объединение польских ученых-эмигрантов, объявило конкурс на оценку работ Вронского. На конкурс поступила только одна работа, по-видимому, потому, что, кроме скромного числа поклонников Вронского, его труды не находили последователей, вероятно, из-за отсутствия ссылок на современных ему математиков и весьма трудного языка. (Вронский применял огромное количество обобщений и объединял в одно целое понятия философского и математического порядка.) Кроме того, изложение Вронского отличалось неприятным безусловным культом собственных тезисов, исключавшим всякую критику. Необходимо отметить, что за свою жизнь Вронский не встретился с критиком, который, наряду с указанием на неясности и неточности в его трудах, упомянул бы о его действительно гениальных идеях.

Гёне-Вронский прочно вошел в историю разработки теории дифференциальных уравнений. Он первым обратил внимание на функциональный определитель, имеющий большое значение в теории обыкновенных линейных дифференциальных уравнений. Этот функциональный определитель, введенный им впервые в 1812 году, так и называется определителем Вронского, или «вронскианом».

Вронский прожил долгую (умер он 9 августа 1853 года в Париже), но трудную жизнь. Было даже такое время, когда ему для спасения больной жены пришлось продать свои ботинки. Незадолго до смерти он еще был полон планов дальнейших работ, но за несколько дней до конца, в предчувствии его приближения, сказал: «Боже..., мне еще столько нужно сказать!».



КАРЛ ФРИДРИХ ГАУСС

(1777 – 1855)

Ценность научного творчества бывает различна. Для общего прогресса человечества наиболее ценным является научное творчество, устанавливающее новые пути исследований, новые направления, по которым идут последователи. К числу ученых-новаторов принадлежит гениальный немецкий математик Карл Фридрих Гаусс.

Гаусс родился 30 апреля 1777 года в Брауншвейге в семье поденного рабочего. Учился в Геттингенском университете в 1795 – 1798 годах. Докторскую степень получил в Хельмштадте в 1799 году за точное доказательство основной теоремы алгебры о том, что всякое алгебраическое уравнение n -й степени с действительными коэффициентами имеет точно n корней. Позднее (в 1814 и 1850 годах) Гаусс привел еще три доказательства этой теоремы. Следует отметить, что доказательства этой теоремы, данные до Гаусса Д'Аламбером, Эйлером, Лагранжем и Лапласом, не были точными.

В 1807 году Гаусс был назначен заведующим кафедрой математики Геттингенского университета и одновременно директором астрономической обсерватории в Геттингене. На этом посту он оставался вплоть до самой смерти. Скончался Гаусс 22 февраля 1855 года.

Гауссу пришлось жить в очень бурный исторический период (наполеоновские войны, французская буржуазная революция, весна народов 1848 года), но он пребывал вне социальных и политических событий своего времени. Гаусс полностью посвятил себя научному труду, результаты которого в математике и других родственных отраслях знаний стали тем, чем были события его эпохи в историческом развитии европейских наций.

Основная черта научных работ Гаусса – это их исключительная разносторонность. Он занимался высшей алгеброй, теорией чисел, дифференциальной геометрией, теорией вероятностей (открыл так называемое нормальное распределение Гаусса), теорией притяжения, теорией электричества и магнетизма, вопросами капиллярности, геодезией и астрономией. Во всех упомянутых областях знаний Гаусс сделал свои оригинальные открытия.

Первый крупный труд Гаусса «Арифметические исследования» содержит его работы по теории чисел и высшей алгебре, включая интересную теорию уравнения деления круга, т.е. уравнения $x^n = 1$, которая давала возможность с помощью этого уравнения строить правильные многоугольники. Гаусс показал, что если число сторон правильного многоугольника есть простое число вида $2^{2^n} - 1$, то такое построение возможно. Сам Гаусс построил правильный семнадцатиугольник. Он придавал этому решению очень большое значение, что видно из его завещания начертать на его могильном памятнике чертеж пра-

вильного семнадцатиугольника, вписанного в окружность. Желание Гаусса исполнено: на памятнике, поставленном в его честь в Брауншвейге, изображен семнадцатиугольник. Весьма ценные результаты получил Гаусс в дифференциальной геометрии, рядах (гипергеометрический ряд Гаусса) и в эллиптических функциях. Гаусс считается также одним из создателей неевклидовой геометрии. Он применил теорию комплексных чисел при решении различных задач. Сам термин «комплексное число» принадлежит Гауссу, причем плоскость комплексных чисел так и называют плоскостью Гаусса.

Заслуги Гаусса в астрономии не меньше, чем в математике. Основной его труд по астрономии «Теория движения небесных тел» содержит способ определения орбит планет на основе наблюдений; Гаусс с успехом применял способ наименьших квадратов. Свои неожиданные астрономические знания Гаусс использовал при нахождении планетоиды Цереры, открытой астрономом Пиацци из Палермо и затем утерянной. То, что не удалось другим математикам, удалось Гауссу: на основе полученных результатов наблюдений Гаусс рассчитал орбиту планетоиды, что позволило снова найти утерянный объект. Велики заслуги Гаусса и в физике, что, в частности, отражено в названии единицы магнитной индукции «Тгаусс». Кроме теоретических работ Гаусса по физике, следует отметить изобретенные им физические приборы. В области физики Гаусс сотрудничал с В. Вебером. Результатом этого сотрудничества явилось изобретение в 1833 году первого в Германии электромагнитного телеграфа.

Однако Гаусс, несмотря на великолепные достижения в различных областях науки, был прежде всего математиком. В его жизни можно отметить периоды, когда он работал в других областях науки, но и тогда он не забывал о математике как теоретической, так и прикладной. В авторских указателях фамилий, обыкновенно помещаемых в конце книг на математические темы, рядом с фамилией Гаусса можно найти большое количество номеров страниц, потому что Гаусс в своих работах коснулся многих вопросов, рассматриваемых на страницах книг. Теорему Гаусса-Остроградского знает любой студент технического вуза. Современники Гаусса называли его «*princeps mathematicorum*» (король математиков). В заключение приведем анекдот из детских лет Гаусса о сумме натуральных чисел, рассказанный Еленским, автором книг «По следам Пифагора» и «Пилавати».

«Как только Карлуше исполнилось 7 лет, его по обычаю отдали в начальную школу. Учителем математики в школе был пожилой человек, известный своей суровостью. Иногда, желая облегчить себе работу по просмотру тетрадей учеников старших классов, он задавал младшим школьникам трудную письменную работу, с тем, чтобы в классе царил полное молчание. Ученик, закончивший работу, обязан был отнести её учителю и положить на кафедре. Однажды учитель дал ученикам следующую задачу: «Найти сумму всех чисел от одного до сорока». Он был уверен, что ученики большую часть урока будут заняты расчетами. Каково же было удивление старого педагога, когда сразу же после написания задачи на доске он услышал веселый крик: «У меня готово!» и

учителю была вручена тетрадь ученика Карла Гаусса. Разгневанный учитель, считая, что Гаусс вздумал подшутить над ним, пробормотал, не прерывая своей работы: «Ну, подожди-ка, шалунишка, я тебя скоро отучу от подобных шуток!» Стоит ли говорить, что решение Гаусса:

$$1 + 2 + \dots + 19 + 20$$

$$40 + 39 + \dots + 22 + 21$$

$$41 + 41 + \dots + 41 + 41, \text{ т.е. } 41 \cdot 20 = 820$$

– как читатель легко заметит – было правильным, и суровый учитель не только не отучил Гаусса от шуток, но занялся серьезно его образованием».



ДАВИД ГИЛЬБЕРТ

(1862 – 1943)

Наряду с французским математиком Пуанкаре, Давид Гильберт принадлежит к числу самых выдающихся математиков своего времени. Родился он 23 января 1862 года в Кенигсберге. За достижения в области теории алгебраических инвариантов ему присвоено звание члена Академии. В 1895 году Гильберт был назначен профессором в Геттингене, где он и работал до конца жизни. Характерной чертой творчества Гильберта была многолетняя и упорная работа над одним избранным узким вопросом, и надо сказать, что этот великолепный мастер достигал больших успехов в любой области математики. После первого успеха в области теории инвариантов (1885 – 1893) Гильберт занялся теорией алгебраических чисел (1893 – 1898). В итоге возник весьма оригинальный труд. После этого Гильберт заинтересовался основами геометрии. Оказалось, что система аксиом как Евклида, так и Лобачевского, не была полной и достаточной, геометрия этих ученых содержала много лишних аксиом, вытекающих из предыдущих. Увлечение Гильберта геометрией привело к созданию полной, независимой системы аксиом. Гильберт опубликовал эту систему в труде «Основы геометрии» (1898). Создание системы аксиом считается крупнейшим достижением этого математика.

В элементарной геометрии, которая преподается в средней школе, мы пользуемся не системой Гильберта, а системой, близкой к Евклидовой, так как точное применение системы Гильберта было бы слишком трудным. В 1900 году на Международном конгрессе математиков в Париже Гильберт поставил 23 проблемы, которые до сегодняшнего дня занимают математиков всего мира. В течение нескольких лет Гильберт работал над разрешением ряда проблем вариационного исчисления и над теорией дифференциальных уравнений. В истории теоремы Варинга, которая гласит, что для всякого натурального числа n можно подобрать такое натуральное число k , что любое натуральное число N будет суммой n -х степеней k целых положительных чисел, последнее слово осталось снова за Гильбертом. Эта теорема была поставлена в 1782 году и по своему содержанию является элементарной. Несмотря на это, она не была доказана в течение ста с лишним лет. И только в 1909 году Давид Гильберт дал доказательство, используя труднейшие выводы высшей математики.

Некоторые достижения математиков в теории интегральных уравнений вызвали интерес Гильберта к этому вопросу, что привело его к новому, отмеченному на страницах истории математики, достижению.

Дополнительным результатом этой работы явилось создание и введение в математику понятия, известного теперь как «пространство Гильберта». Оставив интегральные уравнения, Гильберт вернулся к проблеме, которую считал важ-

нейшей для современной математики, а именно к созданию логических основ математики. Этой проблеме Гильберт посвятил последние годы своей жизни, доказав тем самым, что он до конца остался на трудном посту и упорно, шаг за шагом, стремился к развитию и унификации прекрасной науки: математики. Умер Гильберт 14 февраля 1943 года в Геттингене.



РЕНЕ ДЕКАРТ

(1596 – 1650)

Рене Декарт больше известен как великий философ, чем как математик. Но именно он был пионером современной математики, и его заслуги в этой области столь велики, что он по справедливости входит в число великих математиков современности. О жизни Декарта, известного также под латинизированным именем Картезия (отсюда картезианство), мы знаем немного.

Родился Декарт во Франции, в небольшом городке Лаз, в Департаменте Турень. После окончания иезуитского колледжа для сыновей аристократических семейств он по примеру своего брата стал изучать правоведение. В 22-летнем возрасте уехал из Франции и в качестве офицера-добровольца служил в войсках разных военачальников, участвовавших в тринадцатилетней войне. Во время войны ему приходилось бывать в Австрии, Венгрии и на территории нынешней Чехословакии. Декарт в своем философском учении развивал идею о всемогуществе человеческого разума, и поэтому преследовался католической церковью. Желая найти убежище для спокойной работы по философии и математике, которыми он интересовался с детства, Декарт в 1629 году поселился в Голландии, где прожил почти до конца жизни. В одном из писем Декарт так описывал свое пребывание в Голландии: «Здесь все, кроме меня, так заняты своими делами и доходами, что можно прожить всю жизнь и никто вами не интересуется. В какой другой стране можно пользоваться большей свободой, где можно спать с большим спокойствием, чем здесь, где яд, предательство, клевета распространены значительно меньше, чем в других странах...»

Все крупные произведения Декарта по философии, математике, физике, космологии и физиологии написаны им в Голландии.

Математические труды Декарта собраны в его книге «Геометрия» (1637) В «Геометрии» Декарт дал основы аналитической геометрии и алгебры. Декарт был первый, кто ввел в математику понятие переменной функции. Он обратил внимание на то, что кривая на плоскости характеризуется уравнением, обладающим тем свойством, что координаты любой точки, лежащей на этой линии, удовлетворяют данному уравнению. Он разделил кривые, заданные алгебраическим уравнением, на классы в зависимости от наибольшей степени неизвестной величины в уравнении. Декарт ввел в математику знаки плюс и минус для обозначения положительных и отрицательных величин, обозначение степени $(x \cdot x) = x^2$ и знак ∞ для обозначения бесконечно большой величины. Для переменных и неизвестных величин Декарт принял обозначения $x, y, z, \dots$, а для величин известных и постоянных – $a, b, c, \dots$; как известно, эти обозначения применяются в математике до сегодняшнего дня.

Декарт положил начало исследованию алгебраических уравнений. В частности, он установил, что число действительных и мнимых корней алгебраического уравнения равняется степени неизвестного. Это является важнейшей теоремой алгебры, доказанной значительно позднее Гауссом. Известно также правило Декарта относительно числа положительных корней уравнения с действительными коэффициентами, согласно которому это число равно (или меньше на четное число) количеству изменения знаков в последовательности коэффициентов уравнения

$$a_0 \cdot x^n + a_1 \cdot x^{(n-1)} + \dots + a_{(n-1)} \cdot x + a_n = 0 \quad (a_0 \neq 0, a_n \neq 0).$$

Позднее, на основе математических достижений Декарта, благодаря Лейбницу и Ньютону, были разработаны принципы дифференциального исчисления. В физике Декарт открыл законы отражения и деформации волн и объяснил причины появления радуги. Несмотря на то, что в области аналитической геометрии Декарт продвинулся не очень далеко, его труды оказали решающее влияние на дальнейшее развитие математики. На протяжении 150 лет математика развивалась путями, предначертанными Декартом.



ПЕТЕР ГУСТАВ ЛЕЖЁН ДИРИХЛЕ

(1805 – 1859)

Петер Густав Дирихле (правильнее Дерихле), крупнейший немецкий математик, родился 13 февраля 1805 года в Дюрене, Рейнской провинции. В 1822 году он переехал в Париж, где поселился в доме генерала Фуа. Здесь ему представился удобный случай познакомиться со многими знаменитыми учеными, философами и математиками. В это же время он посещал лекции в Coliege de France и глубоко изучил труд Гаусса: «Disquisitiones arithmeticae», что дало направление его исследовательским устремлениям. В 1826 году Дирихле возвратился в Германию, где получил должность приват-доцента в Бреславльском университете (ныне Вроцлавском), а потом переехал в Берлин. Здесь он в 1829 году сначала стал приват-доцентом, затем, в 1831 году, ординарным профессором в университете. Одновременно он стал преподавателем военного училища. В 1855 году Дирихле был приглашен в Геттингенский университет в качестве продолжателя Гаусса. Оригинальное творчество Дирихле касается, в основном, теории чисел, теории рядов, интегрального исчисления и некоторых проблем математической физики. В 1825 году Дирихле написал труд «Memoire sur l'impossibilite», который, будучи представлен Парижской академии, обратил на него внимание ученых и обеспечил ему славу прекрасного математика. В этой работе Дирихле рассмотрел случай так называемой великой теоремы Ферма для $n=5$ (Эйлер и Лагранж рассматривали случай $n=3$ и $n=4$). После этого Дирихле дал доказательство теоремы Гаусса для двуквадратичных остатков. Дирихле показал большую роль анализа и теории аналитических функций для решения проблем теории чисел. Известна доказанная им теорема о существовании бесконечно большого числа простых чисел во всякой бесконечной арифметической прогрессии из целых чисел, первый член и разность которой – числа взаимно простые. До Дирихле эта проблема представляла для математиков непреодолимые трудности. Дирихле первым дал точное доказательство сходимости рядов Фурье, известное повсеместно как признак Дирихле, а в вариационном исчислении привел так называемый принцип Дирихле. Эти работы дали повод другим математикам, например Риману и Кантору, углубить исследования, что привело их к новым открытиям. Значительные работы Дирихле посвящены механике и математической физике.

Свои исследования и трактаты Дирихле печатал в математическом журнале Крепла и в трудах Академии. Он не написал крупного произведения, но его научное наследие и его лекции значительно повлияли на развитие математических знаний в Германии. Дирихле умер 5 мая 1859 года в Геттингене. После смерти ученого его лекции по теории чисел в обработке Дедекинда стали классическим трудом.



СОФИЯ ВАСИЛЬЕВНА КОВАЛЕВСКАЯ

(1850 – 1891)

В числе сотни наиболее выдающихся математиков последних веков почетное место принадлежит женщине. Это – София Ковалевская, урожденная Корвин-Круковская. София Ковалевская родилась 15 января 1850 года.

Женщины царской России были лишены права получать высшее образование, они не могли поступать в высшие учебные заведения. Поэтому Ковалевская встретила на своем пути к научной карьере большие препятствия. Она не смогла поступить в Московский университет и начала изучать математику частным образом, так как именно в этой отрасли знаний уже с ранних лет отличалась большими способностями. После этого она выехала в Германию и только там, в Гейдельбергском университете, могла слушать курс математики и физики. Позднее она переехала в Берлин и там обратила на себя внимание выдающегося математика Вейерштрасса. Восхищенный её способностями, Вейерштрасс стал давать ей частные уроки. Они совместно обсуждали многие научные проблемы, главным образом касающиеся неевклидовой геометрии. Позже Вейерштрасс говорил, что у него было очень немного учеников, которые могли бы равняться способностями, прилежанием и любовью к науке с Ковалевской. В это время Ковалевская написала три работы, из которых каждая, по мнению Вейерштрасса, была вполне достаточна для присуждения Ковалевской докторской степени. Это работы: «Некоторые данные к теории уравнений в частных производных», «Некоторые данные и замечания к исследованию Лапласа о форме кольца Сатурна» и «О приведении некоторых Абелевых интегралов к интегралам эллиптическим». За эти работы Ковалевской было присвоено звание доктора философии с отличием. Еще и теперь результаты первого из этих трудов приводятся в учебниках математического анализа под наименованием задачи Коши-Ковалевской. Она упростила доказательство Коши и придала теореме окончательную форму, и, что важнее всего, обобщила результат на систему уравнений.

После возвращения в Россию Ковалевская поддерживала оживленные контакты с учеными. Чебышевым, Менделеевым, Столетовым и шведским основателем «Acta Mathematica» математиком Миттаг-Лефлером, благодаря которому в 1883 году была назначена заведующей кафедрой математики Стокгольмского университета. Она была очень рада этому назначению, потому что, по её собственным словам, «функции профессора содержат в себе что-то благодарное. Не говоря уже о большом значении, какое мне лично придают обязанности профессора, я была бы очень счастлива, если бы могла пробить новую дорогу для женщин...» Назначение женщины профессором вызвало крайнее

возмущение среди консервативных университетских кругов. За Ковалевской отрицали какие-либо научные заслуги и часто клеветали на неё. Но были у неё и друзья. В одной из демократических газет была помещена статья, в которой, в частности, было написано: «На нашу долю выпала большая честь сообщить о приезде в наш город не какого-либо обыкновенного принца крови. Нам оказала честь своим посещением принцесса науки, госпожа Ковалевская, которая будет первым профессором-женщиной в Швеции». В 1888 году она написала свой основной научный труд «Задача о вращении твердого тела вокруг неподвижной точки». За эту работу Ковалевская была удостоена первой премии на конкурсе, объявленном Парижской академией наук. Вскоре Ковалевская получила награду от Шведской академии наук за свои научные достижения. Она сотрудничала с такими великими французскими математиками, как Пуанкаре, Эрмит. Ковалевская получила известность во всех существовавших тогда научных кругах. Пожалуй, не лишним будет упомянуть о её больших способностях к языкам и литературе (она владела пятью языками, писала научно-популярные статьи в газеты, театральные рецензии, рассказы, написала несколько романов автобиографического содержания, в частности роман «Нигилистка», запрещенный цензурой; роман был напечатан лишь в 1928 году).

Ковалевская умерла от воспаления легких в полном расцвете сил и славы 10 февраля 1891 года в Стокгольме.



ОГЮСТЕН ЛУИ КОШИ

(1789 – 1857)

Коши родился 21 августа 1789 года в Париже в семье юриста. День его рождения почти совпал с началом Великой французской буржуазной революции. После окончания в 1807 году Политехнической школы и в 1810 году Школы строительства дорог и мостов он, по назначению правительства, работал в качестве инженера-строителя морских портов. По-видимому, тогда он посвящал много времени королеве наук – математике, так как уже в 1811 году представил Академии наук в Париже работу по теории многогранников, обратившую на него внимание парижских учёных. Коши довольно быстро приобрел известность и в 1816 году получил высокое ученое звание члена Академии наук. Он преподавал в College de France, в Сорбонне, в Ecole Polytechnique. Членом Академии он стал на основе конкурсной работы «О распространении волн на поверхности жидкости».

В 1830 году, после июльской революции, Коши отказался присягнуть новому правительству Франции и уехал вместе с семьей за границу. Сначала он жил в Швейцарии, потом два года в Турине, где возглавлял кафедру математической физики, которую сам же основал. В 1833 году выехал в Прагу и в течение пяти лет занимался там воспитанием сына императора Карла X, за что получил в награду титул барона. После возвращения во Францию в 1838 году ему вернули все прежние звания. По отношению к Коши было сделано исключение: ему разрешили вести педагогическую деятельность без присяги на верность правительству.

Коши написал свыше 800 трудов. Этому благоприятствовали не только трудолюбие Коши и гениальность его ума, но и внимание к его работам со стороны современников.

В богатом научном наследии Коши есть работы различного типа из разных разделов математики. В них он представил результаты своих собственных исследований, отчеты о работах, присылаемых в Академию, и результаты дидактической деятельности – превосходные учебники математического анализа, которые стали образцом научного мышления для последующих поколений математиков.

Работы Коши посвящены анализу, геометрии, теории чисел и математической физике. Например, в алгебре Коши принадлежит прекрасное доказательство основной теоремы; в теории чисел Коши доказал, что любое натуральное число является n -угловым числом, или может быть представлено суммой не больше чем n чисел n -угловых; в математической физике ему принадлежит обобщенное математическое понятие деформации и упругого напряже-

ния. В работах по теории упругости он рассматривал тело как сплошную среду и оперировал напряжением и деформацией, относимой к каждой точке.

Особым и доминирующим разделом научного творчества Коши является математический анализ, в частности, проблемы функций комплексного переменного и теория дифференциальных уравнений. Заслугой Коши является собрание и уточнение богатого наследия математиков XVIII столетия. Он последовательно пользовался понятием предела и введенным им самим понятием непрерывности функций. Еще и теперь математики всего мира пользуются такими понятиями, как «предел функции Коши», «непрерывность функции по Коши», «произведение рядов по Коши», «признак сходимости рядов по Коши», «метод Коши» (интегрирование дифференциальных уравнений), «остаток Коши», «теорема Коши (и Пеано) о существовании решения дифференциального уравнения «Коши проблемы», т. е. одной из так называемых «краевых задач», «Коши интегральная теорема», «интеграл Коши», «неравенство Коши».

В истории дифференциального и интегрального исчисления (в их широком понимании) и вообще в истории математики Коши занял выдающееся место. В работах по оптике он дал математическую разработку теории Френеля и теории дисперсии.



ЖОЗЕФ ЛУИ ЛАГРАНЖ

(1736 – 1813)

Жозеф Лагранж родился в 1736 году в Турине, в семье банковского чиновника. Будучи самым младшим сыном в многочисленной семье, он вынужден был сделать все, чтобы как можно скорее начать самостоятельную жизнь. Сначала он интересовался филологией, но когда в его руки случайно попал трактат по математической оптике, он почувствовал свое настоящее призвание. Он целиком посвятил себя изучению математики и в семнадцатилетнем возрасте был назначен преподавателем математики в Королевской артиллерийской школе в Турине. Несмотря на то, что Лагранж был моложе многих своих учеников, он пользовался славой прекрасного преподавателя.

В 1757 году Лагранж вместе со своими бывшими учениками основал Туринскую академию наук, которая первоначально носила характер частного научного общества. В этой Академии Лагранж был председателем физико-математической секции. В 1759 году вышел в свет первый том трудов Академии «Actes de la Societe privee». В этом томе были напечатаны многие труды двадцатитрехлетнего ученого, в основном по вариационному исчислению в его связи с механикой. На основе работ Эйлера, Лагранж дал основные понятия вариационного исчисления и предложил метод решения вариационных проблем. После ознакомления с этими трудами Эйлер в письме к Лагранжу написал так: «...Ваше решение изопериметрических проблем безукоризненно, и я рад, что тема, которой я давно занимаюсь, доведена Вами до близкого конца...»

В том же 1759 году Лагранж был избран членом Берлинской академии наук. Вследствие переутомления Лагранж в двадцатипятилетнем возрасте тяжело заболел.

В 1766 году ученый переехал в Берлин, где получил должность председателя Берлинской академии наук, в которой он оставался бессменно на протяжении двадцати лет. Это был необыкновенно плодотворный период жизни Лагранжа. Он почти ежемесячно опубликовывал новую работу. Из его берлинских работ важнейшей является теория частных дифференциальных уравнений. В 1772 году Лагранж был избран членом Академии наук в Париже, куда переехал в 1787 году и стал преподавать в высших учебных заведениях. В 1788 году в классическом трактате «Аналитическая механика» Лагранж расширил основы статики и механики, установив «общую формулу», известную теперь как принцип возможных перемещений. На протяжении своей жизни Лагранж опубликовал много трудов по математическому анализу (теорема Лагранжа и вытекающие из неё выводы стали основой для исследования так называемого хода изменемости функций; остаточный член формулы Тейлора). В своих трудах Лагранж коснулся теории чисел, алгебры, интерполяции (интерполяционная формула Лагранжа), математической картографии и астрономии. Кроме теоре-

тических работ, следует упомянуть участие Лагранжа в трудах комиссии по выбору новой единицы длины (метр) и разработке на этой основе метрической системы мер.

Умер Лагранж в 1813 году после длительной болезни, в результате полного истощения сил. Выдающийся ученый Лаплас в речи над могилой Лагранжа дал такую характеристику деятельности покойного: «...среди тех, кто самым эффективным образом раздвинул пределы наших знаний, Ньютон и Лагранж в самой высокой степени владели счастливым искусством открывания новых данных, представляющих собой существо знаний...»



ПЬЕР СИМОН ЛАПЛАС

(1749 – 1827)

Лаплас родился 23 марта 1749 года в Бомон в Нормандии. Учился в школе монашеского ордена бенедиктинцев. В 1766 году приехал в Париж, где благодаря помощи Д'Аламбера получил назначение профессором в Парижскую военную школу. В 1779 – 1784 годах Лаплас совместно с Лавуазье занимался физическими вопросами, в частности теплотой плавления тел.

Великая французская революция прервала работы Лапласа в этой области. Ему поручили реорганизацию новых высших учебных заведений – *Ecole Normale* и *Ecole Polytechnique*.

В 1790 году Лаплас был избран председателем Палаты мер и весов и занялся внедрением в жизнь метрической системы мер. В 1795 году Лаплас был назначен руководителем Бюро Долгот, которое занималось измерением длины земного меридиана. В 1799 – 1823 годах Лаплас разработал и опубликовал пятитомный труд «Механика неба» и «Аналитическая теория вероятностей» (1812). В обоих трудах Лаплас изложил не только свои работы, но и все имевшиеся на тот момент сведения в затронутых областях знания. Кроме астрономии, Лаплас занимался некоторыми вопросами математики и математической физики. В этой области достаточно указать на его труды по интегрированию уравнений с частными производными и на теорему о представлении определителей суммой произведений дополнительных миноров на последовательные элементы произвольного столбца или строки, а также на интегральное преобразование, получившее его имя. Несмотря на то, что до Лапласа теорией вероятностей занимались, в частности, Ферма и Бернулли, именно Лаплас является в значительной мере автором математической теории вероятностей, причем ему принадлежит усовершенствование методов доказательства. Он разработал теорию ошибок и способ наименьших квадратов, доказал теорему о пределах, получившую его имя, которая до сих пор имеет большое значение в теории вероятностей. Его «Аналитическая теория вероятностей» выдержала три прижизненных издания.

Всю свою жизнь Лаплас интересовался астрономией, математикой и физикой. В 1806 году опубликовал труд по теории капиллярности, в 1809 году вывел формулу для скорости распространения звука в воздухе и барометрическую формулу для определения плотности воздуха, в зависимости от высоты над уровнем моря, выдвинул теорию истечения света. Деятельность его в области физики привела к значительному развитию экспериментальной физики.

В астрономии Лаплас доказал, что закон всемирного тяготения полностью объясняет движение планет. В 1780 году он представил труд, в котором разработал методы расчета орбит небесных тел, а позднее выдвинул гипотезу, что

кольца Сатурна не однородны, а сам Сатурн должен быть сжат у полюсов. В 1789 году разработал теорию движения спутников Юпитера, полностью подтвержденную позднейшими наблюдениями. В 1787 году открыл причины ускорения движения Луны и определил величину сжатия Земли. В 1788 году опубликовал труд по динамической теории приливов. Все свои научные открытия в этой области знания он опубликовал в книге «Механика неба».

Следует отметить, что космогоническая теория Лапласа, так называемая гипотеза Канта-Лапласа, имеет также огромное философское значение. Гипотеза была опубликована Лапласом в труде «Изложение системы мира» (1796).

По своим философским взглядам Лаплас склонялся к материализму. Несмотря на то, что первоначальное образование ученый получил в монастырской школе, он всю жизнь был атеистом. Что касается политических убеждений, то они менялись у него в зависимости от форм правления во Франции. В период революции Лаплас был удостоен многих почетных наград. Во время Консулата состоял министром внутренних дел, а позднее стал сенатором. В 1811 году Наполеон присвоил ему звание графа де Лаплас, за что ученый посвятил императору третий том своего «Трактата о небесной механике». Говорят, что на вопрос Наполеона, почему в его «Трактате» нет упоминания о Боге, Лаплас ответил: «Ваше величество, эта гипотеза оказалась ненужной». Благорасположение императора не помешало Лапласу проголосовать в 1814 году за низложение Наполеона. Знаменательно и то, что после 1814 года третий том своего «Трактата о небесной механике» он посвятил Людовiku XVIII. В период реставрации Лаплас получил от Бурбонов титул маркиза и звание пэра Франции.

Умер Лаплас 5 марта 1827 года в Париже.



ДЖОН ФОН НЕЙМАНН

(1903 – 1957)

Джон фон Нейманн, инициатор строительства современных вычислительных машин, родился 28 декабря 1903 года в Будапеште. Он отличался чрезвычайной памятью и в ранней молодости проявлял необыкновенные способности и любовь к точным наукам. Учился в Берлинском университете, где изучал сначала химию, потом математику. Закончил также Technische Hochschule в Цюрихе и Будапештский университет. Будучи еще молодым, а именно в 1927 году, Джон фон Нейманн стал приват-доцентом в Берлинском университете, затем читал лекции в Гамбургском университете. В 1930 году по приглашению Принстонского университета выехал в Соединенные Штаты Америки, где и остался навсегда.

Нейманна интересовали многие разделы математики, но его научные труды первоначально носили теоретический характер. Он занимался теорией функции действительного переменного, математической логикой, теорией размерностей, непрерывной геометрией, общей топологией, эргодической теорией, проблемами колец операторов, теорией квантов. Работая в области теории вероятностей, он интересовался развитием теории игр и экономической математики. В конце тридцатых годов увлекся проблемами теоретической гидродинамики и много времени посвятил решению частных дифференциальных уравнений аналитическим методом.

В период нарастания угрозы мировой войны Нейманн начал работу по применению математики к физике, в частности, он много времени посвятил вопросу взаимодействия ударных волн, что имело большое военное значение. Эта проблема потребовала огромной вычислительной работы, что и стало основной причиной интереса Нейманна к вычислительным машинам. Он стал ведущим знатоком и пропагандистом использования электронных вычислительных машин при решении различных научных проблем. Нейманн модернизировал и усовершенствовал машину ЭНИАК, построенную в Филадельфии для Ballistik Research Laboratories. Он работал в Вашингтоне, Лос-Аламосе и в других городах Соединенных Штатов, участвовал во многих научных конференциях. После войны сотрудничал с группой математиков и инженеров и построил экспериментальный электронный вычислитель в Institute for Advanced Study, под наименованием ДЖАНИАК. Следует заметить, что в основу проекта была положена особенность работы человеческого мозга. Нейманн специально изучил неврологию и психиатрию и пришел к убеждению, что электронные машины могут строиться, как упрощенные модели человеческого мозга. Перегруженность работой и недостаток времени привели к тому, что ученый вынужден был писать

свои научные труды только по ночам или ранним утром. Нейманн пытался решить ряд проблем по метеорологии с помощью вычислительных машин, которыми он пользовался также при решении многих задач ядерной физики.

Нейманн тесно сотрудничал с лабораториями Комиссии по делам атомной энергии, а в 1952 году был назначен членом Общего консультативного комитета АЕС. После принятия 15 марта 1955 года присяги Нейманн был назначен членом Комиссии по делам атомной энергии и в мае этого же года вместе со своей женой Кларой поселился в Вашингтоне. В августе у него появились первые признаки рака кости.

В начале 1955 года руководители Фонда Силлимана пригласили Нейманна прочесть несколько лекций в Иельском университете в весеннем семестре 1956 года. В Соединенных Штатах Америки приглашение подобного рода считается большой привилегией потому, что получить его могут только ученые, пользующиеся мировой известностью. По традиции лекции издаются в виде книги на средства Иельского университета. Нейманн со всем усердием взялся за разработку темы лекции и написал свой последний труд «Вычислительная машина и человеческий мозг». С января 1956 года он был вынужден пользоваться инвалидной коляской, но несмотря на это работал в своей канцелярии и участвовал в научных собраниях. В начале апреля Нейманна поместили в больницу Уолтера Рида, где он, несмотря на упадок сил, продолжал работу над рукописью, которую, к сожалению, не закончил.

Нейманн умер 8 февраля 1957 года. Иельский университет в честь знаменитого ученого издал его рукопись в серии «Лекции Силлимана».



ИСААК НЬЮТОН

(1642 – 1727)

О Ньютоне, пожалуй, слышали все, и все знают, что он был великим физиком. Действительно, значительное развитие современной физики начинается с Ньютона. О Ньютоне как математике знают немногие, несмотря на то, что достижения его в этой области так же эпохальны, как и в физике.

Ньютон родился 4 января 1642 года в семье бедного фермера в Вулсторе, в 75 километрах от Кембриджа, в Англии. После окончания школы он поступил в Тринити колледж (один из колледжей Кембриджского университета). Там он получил Степень магистра (1668). Вскоре Ньютон получил кафедру математики и физики в Кембриджском университете, которой руководил тридцать два года. Первые лекции он посвятил оптике. По словам самого Ньютона, наиболее плодотворными в его научной работе были 1665 и 1666 годы.

В истории науки, пожалуй, не найти достижений, достойных сравнения с трудами Ньютона в эти два золотых года. Он первый (одновременно с Лейбницем) создал основы дифференциального и интегрального исчисления, начал работу над своим крупным произведением об оптике «*New Theory about Light and Colours*» («Новая теория света и цветов»), создал основы теории всемирного тяготения. Ньютон первый указал на то, что луч белого света после прохождения через призму расщепляется на лучи разных цветов. Его труд по оптике вызвал горячие споры в научном мире о природе света. Все эти гениальные идеи зародились в уме Ньютона во время его пребывания в родной деревне, куда он выехал, спасаясь от эпидемии, свирепствовавшей в Кембридже. Крупнейшим произведением Ньютона было «*Philosophiae naturalis principia mathematica*» («Математические начала натуральной философии»), вышедшее в 1687 году. Ньютон привел в нем определение трех основных принципов классической механики и открытый им закон тяготения, на основе которого он разработал теорию движения планет и объяснил много других проблем астрономии (в частности, причину морских приливов и отливов на Земле, вызываемых притяжением Луны). Все эти достижения в физике были бы невозможны без одновременного развития математических методов. Мы уже знаем, что Ньютон был одним из создателей дифференциального и интегрального исчисления, принципы которых были опубликованы в двух его трудах «*De quadratura curvarum*» («Рассуждения о квадратуре кривых», 1704) и «*The Method of Fluxions and Infinite Series*» («Метод флюксии и бесконечных рядов», 1736). В этих трудах Ньютон разработал основы математического анализа, т. е. дал решение таких вопросов, как нахождение экстремумов функций, точек перегиба, уравнений касательных к кривым, вычисление длины кривых, площадей, ограниченных кривыми, и дал методы решения простых дифференциальных урав-

нений. В частности, Ньютон создал метод приближительного решения уравнений, так называемый метод касательных (метод Ньютона). Кроме того, здесь же Ньютон дал свою известную формулу (бином Ньютона), в которой приводится решение выражения $(a+b)^n \cdot (a+b)^n$ при любой натуральной степени n , а также много других математических правил и теорий. В 1672 году Ньютон был избран членом Королевского научного общества в Лондоне (Royal Society), позднее стал его председателем. В девяностых годах Ньютон серьезно заболел, что вызвало большую тревогу в научном мире. Об этом свидетельствует, например, обмен письмами таких выдающихся математиков, как Лейбниц и Гюйгенс. Однако болезнь вскоре прошла, и в том же году Ньютон был избран членом Академии наук (Academic des Sciences) в Париже. Спустя некоторое время Иоганн Бернулли разослал наиболее выдающимся математикам две задачи, требующие решения, и дал для этого срок шесть месяцев. К своему удивлению, очень скоро Бернулли получил анонимное решение. Но он догадался, что автором его мог быть только Ньютон.

К сожалению, довольно трудно определить влияние научных достижений Ньютона на его современников, так как он очень поздно печатал свои труды. Например, закон всемирного тяготения, открытый им в 1665 – 1666 годах, он опубликовал спустя двадцать лет. Трактат «Arithmetica universalis» («Всеобщая арифметика»), содержащий лекции по алгебре, прочитанные им в 1663 – 1683 годах, напечатал только в 1707 году. То же самое получилось с основами дифференциального исчисления, разработанными Ньютоном. Основы были опубликованы уже после выхода в свет подобного труда Лейбница, несмотря на то, что Ньютон написал свою работу на десять лет раньше Лейбница.

Великий ученый умер 31 марта 1727 года.

Как велика была страсть Ньютона к творчеству, его стремление к непрерывной борьбе на научном поприще, свидетельствуют его шутливые слова: «Наука подобна красивой, но сварливой женщине. Если хочешь общаться с ней, надо беспрестанно ссориться».



СИМЕОН ДЕНИ ПУАССОН

(1781 – 1840)

Выдающийся французский физик и математик Симеон Дени Пуассон родился 21 июня 1781 года в Питивье (департамент Луары). В 1798 – 1800 годах обучался в Политехнической школе в Париже и там же начал работать в качестве преподавателя. В 1806 году Пуассон был назначен почетным профессором Политехнической школы, а в 1809 году – профессором Сорбонны. Пуассон в это время прочно вошел в ряды знаменитых математиков. После окончания высшей школы Пуассон интересовался некоторыми астрономическими вопросами, теоретической механикой и применением математики к механике и физике, что позднее привело его к крупным открытиям в теоретической математике. Первый значительный труд Пуассона в двух томах «Трактат механики» появился в печати в 1811 году. Долгое время этот «Трактат» считался лучшим учебником механики. Эта работа Пуассона отличается ясностью и доступностью изложения.

В 1812 году Пуассон был избран членом Парижской академии наук. Спустя год он опубликовал работу под заглавием «Замечания об уравнении теории притяжения», в которой вывел уравнение, получившее впоследствии его имя. Труды Пуассона, а написал он свыше 300 работ, касались различных проблем, например устойчивости солнечной системы, теории упругости, гидромеханики, электростатики и магнетизма. До сегодняшнего дня не потеряли своего значения работы Пуассона по теории определенного интеграла, теории дифференциальных уравнений с частными производными и по теории вероятностей. В 1816 году Пуассон возглавил кафедру теоретической механики Парижского университета, где проработал до 1837 года. В 1826 году был избран членом Петербургской Академии наук.

Из многих трудов, опубликованных Пуассоном, достойны внимания: «Memoire sur la theorie des ondes» («Теория волн», 1826), «Theorie nouvelle de l'action capillaire» («Новая теория капиллярности», 1831), «Theorie du calcul de probabilties» («Теория вероятностей», 1833) и «Theorie mathematique de la chaleur» («Математическая теория распространения тепла», 1835). Будучи сторонником Наполеона, он получил от него титул барона и множество отличий. Несмотря на это, он усердно служил последующим правительствам Франции. Во времена Людовика Филиппа в 1837 году он стал членом палаты пэров.

Умер Пуассон 25 апреля 1840 года в Париже.



ГЕОРГ ФРИДРИХ БЕРНХАРД РИМАН

(1826 – 1866)

Пути развития современной математики в значительной мере были предопределены трудами немецкого ученого XIX века Георга Фридриха Бернхарда Римана.

Риман родился в семье деревенского пастора 17 сентября 1826 года. По настоянию отца он первоначально поступил на теологический факультет Геттингенского университета, но вскоре его наклонности и интерес к точным наукам взяли верх, и молодой студент полностью посвятил себя изучению математики. Риман слушал лекции Гаусса, а потом, уже в Берлинском университете, таких превосходных математиков, как Дирихле, Якоби, Штейнер. Его воспитателем, учителем и другом стал Дирихле. Это весьма благотворно сказалось на развитии творческих способностей Римана. В 1851 году он защитил в Геттингене диссертацию, посвященную теории комплексных функций, и получил степень доктора математики, через три года после представления двух работ: «О возможности представления функции посредством тригонометрического ряда» и «О гипотезах, лежащих в основании геометрии» он был назначен приват-доцентом.

Первая из этих работ была посвящена исследованию условий Дирихле о разложимости функции в ряд Фурье. В этой работе Риман обобщил и дополнил результаты, полученные его учителем. Вторая его работа по геометрии оказала чрезвычайно большое влияние на развитие математических и физических идей. Дело в том, что Риман дал классификацию всех существующих видов геометрии, включая найденные уже неевклидовы геометрии, и показал возможность создания любого числа новых пространств. Эта работа открыла Эйнштейну путь к разработке общей теории относительности.

С течением времени Риман стал преподавать в Геттингенском университете. Говорят, что на его первую лекцию пришло всего лишь восемь человек, а на следующую – и того меньше. Дело в том, что вначале Риман ощущал известные трудности при чтении лекций. Но спустя некоторое время он писал: «...Моя первоначальная застенчивость уже несколько прошла: я привыкаю больше думать о слушателях, чем о себе, и научился читать по их лицам, могли продолжать лекцию или должен еще раз вернуться к рассматриваемой проблеме...»

Застенчивость Римана вскоре совершенно прошла и благодаря тщательной подготовке к лекциям он стал добиваться хороших результатов в обучении студентов. В своих лекциях он часто использовал нигде не опубликованные материалы. После его смерти слушатели тщательно собрали свои записки, и таким образом было создано дополнение к собранию трудов Римана, изданное

почти через сорок лет после его смерти. Некоторое понятие о том, как много сделал Риман для развития математики, может дать перечень методов, теорем и проблем, носящих его имя: теорема Римана-Роха об алгебраических функциях, пространства Римана, интеграл Римана, лемма Римана-Лебега о тригонометрических интегралах, геометрия Римана, гипотеза Римана, Римановы матрицы в теории Абелевых функций, дзета-функция Римана, метод Римана решения частных дифференциальных уравнений гиперболического типа и много, много других. Предложенные Риманом идеи и методы открыли новые пути развития математики и нашли применение в механике и физике. И несмотря на то, что он написал очень немного работ, а напечатал еще меньше, любая из них отличалась огромной важностью и множеством новых идей. К сожалению, туберкулез преждевременно вырвал его из жизни. Риман умер 20 июля 1866 года в местности Селаска, близ Интры в Италии.



ЖАН БАТИСТ ЖОЗЕФ ФУРЬЕ (1768 – 1830)

Жан Батист Жозеф Фурье, великий математик, родился 21 марта 1768 года в семье портного в городе Оксер. Осиротел в возрасте семнадцати лет. С помощью друзей он поступил в военное училище в родном городе, где вскоре отличился благодаря способностям, в особенности математическим. Из-за происхождения и бедности Фурье был лишен возможности выдвинуться на военной службе и поэтому после окончания училища (1784) остался в нем в качестве преподавателя математики, истории и риторики.

В 1796 году Фурье выехал в Париж и возглавил кафедру анализа в Политехнической школе. Сопутствовал Наполеону в 1798 году в его Египетской кампании и был назначен губернатором Нижнего Египта. В течение трех лет был секретарем института, организованного Наполеоном в Каире. Этот институт значительно расширился благодаря Фурье. В 1801 году Фурье возвратился во Францию и в следующем году получил должность префекта департамента Изеры и одновременно титул барона и орден Почетного Легиона. Префектом департамента Изеры Фурье состоял на протяжении четырнадцати лет. В это время он руководил изданием «Description de l'Egypte» и вел трудоемкие и плодотворные исследования по теплопроводности. В 1815 году, после возвращения Наполеона с острова Эльбы, Фурье опубликовал роялистскую прокламацию, за что был лишен должности префекта в Изере и Родане. Начиная с 1817 года он целиком посвятил себя научной работе. Переехал в Париж, где был избран в Академию, но ввиду протеста сторонников Людовика XVIII членом Академии стал лишь год спустя.

С 1827 года занял место Пьера Лапласа в должности ректора Политехнической школы. Умер Фурье 16 мая 1830 года в Париже. Его труды и исследования обеспечили ему мировую известность. Применяемые им методы были совершенно оригинальными, он значительно усовершенствовал теорию уравнений. Ряды, названные его именем (ряды Фурье), сыграли большую роль в математике и часто применяются и поныне. Фурье работал над теорией теплоты и занимался математическим анализом, в особенности теорией функций, интегральным исчислением и дифференциальными уравнениями. Еще в 1807 и в 1811 годах он представил Парижской академии наук свои первые открытия, а в 1822 году издал труд «Аналитическая теория теплоты». Этот труд стал исходной точкой для создания теории тригонометрических рядов и разработки некоторых проблем математического анализа.

Рукопись Фурье по численному методу решения алгебраических уравнений была издана в 1831 году уже после его смерти под заглавием «Анализ определенных уравнений».



ПАФНУТИЙ ЛЬВОВИЧ ЧЕБЫШЕВ (1821 – 1894)

В докладе, прочитанном в декабре 1894 года на заседании Российской Академии наук через несколько дней после смерти Чебышева, академик А.А. Марков сказал: «...двадцать шестого ноября неожиданно умер П.Л. Чебышев. Это невосполнимая потеря для нашей Академии... Имя Чебышева известно в Париже так же, как и в Петербурге. Он уже много лет состоит членом восьми иностранных Академий наук, что свидетельствует о нем как о выдающемся геометре, известном всей Европе...»

В этих словах нет преувеличения, так как Чебышев оказал огромное влияние на развитие мировой науки.

Несмотря на то, что со дня смерти Чебышева прошло уже более века, его труды все еще остаются актуальными. Разработанными темами и способом решения поставленных задач Чебышев создал новую математическую школу, известную в научном мире как школа Чебышева или петербургская математическая школа.

К числу наиболее выдающихся представителей этой школы принадлежали Марков, Ляпунов, Золотарев и многие другие. Математики из школы Чебышева разрабатывали проблемы математического анализа, теории чисел, теории вероятностей.

Юные годы Чебышева мало известны. Он родился 26 мая 1821 года в имении своего отца в селе Окатове Калужской губернии. Семья его принадлежала к старинному дворянскому роду. Получив начальное образование, Чебышев выехал в Москву, где поступил в гимназию. Его учитель математики Погорельский заметил выдающиеся математические способности Чебышева и убедил его поступить на математический факультет. Во время обучения в Московском университете наибольшее влияние на формирование мировоззрения молодого Чебышева оказали профессора Брашман и Ершов.

После четырехлетнего обучения Чебышев закончил университет со степенью кандидата наук. В 1846 году он получил степень магистра и переехал в Петербург, где поступил на должность адъюнкта кафедры теории уравнений Петербургского университета. В 1849 году, после защиты диссертации на тему «Теория сравнений», Чебышев получил научную степень доктора, а три года спустя был назначен экстраординарным профессором; в 1860 году ему присвоили звание ординарного профессора, а в 1872 году – заслуженного профессора. Начиная с 1856 года Чебышев состоял членом Петербургской Академии наук, в 1859 году получил звание действительного члена Академии. В последующие годы Чебышев был избран членом многих иностранных Академий и научных обществ, в частности Парижской академии наук (1860), Королевского

научного общества в Лондоне (1877) и т. д. Чебышев опубликовал много научных работ. Он занимался теорией интерполяции и приближениями функций, применяя при решении этих проблем многочлены, получившие название многочленов Чебышева. Это многочлены вида $T_n(x) = \cos(n \cdot \arccos(x))$ для $n = 0, 1, 2, \dots$ и для $n = 1, 2, 3, \dots$ $U_n(x) = \sin(n \cdot \arccos(x))$. Его труды по теории чисел положили начало деятельности петербургской математической школы. В области теории вероятностей Чебышев опубликовал свой знаменитый труд, касающийся неравенства, до сих пор называемого неравенством Чебышева. В своих дальнейших трудах он открыл закон больших чисел, т.е. одну из так называемых предельных теорем в теории вероятностей.

Большинство проблем, которыми занимался Чебышев, отличается большой практической ценностью. Он решил много задач по механике, сопротивлению материалов и картографии. В частности, он написал работы: «О шестернях», «О построении географических карт» и «О кройке одежды». Казалось бы, работы эти имеют мало общего с так называемой чистой математикой, но Чебышев, работая, например, над вопросом кройки одежды, обосновал практическое применение закона о функциях. Именно практическая ценность его работ принесла Чебышеву мировую славу. Умер Чебышев 8 декабря 1894 года в Петербурге.



ЛЕОНАРД ЭЙЛЕР

(1707 – 1783)

Несомненно, Эйлер принадлежит к числу гениальнейших математиков всех времен. В истории точных наук его имя стоит рядом с именами Ньютона, Декарта, Галилея. Он был не только математиком, но и физиком, и астрономом. Его труды оказали огромное влияние на развитие этих наук. Эйлер родился 4 апреля 1707 года в Базеле в Швейцарии. Он слушал лекции великого математика Иоганна Бернулли, который взялся лично руководить развитием таланта молодого математика. Ученую степень магистра Эйлер получил в возрасте 16 лет. Спустя четыре года он, по приглашению Петербургской Академии наук, выехал в Россию, где стал членом Академии и руководил кафедрой физиологии. С этого времени началось быстрое развитие его научной деятельности. Стимулом для работы над иными отраслями математики становились для Эйлера в значительной мере естественные науки, в особенности механика и техника. Первые его труды касались навигации, но позднее он целиком посвятил себя математике. Журнал Академии («Комментарии») из номера в номер печатал его математические работы. Эйлер был известен необыкновенным трудолюбием, что в конце концов привело его к потере зрения в одном глазу. Это, однако, не помешало его творчеству. Он работал над составлением карты России, написал теорию музыки, позднее издал большой труд о навигации, за который получил 6 тысяч фунтов премии от французского правительства. Мировое признание принесли Эйлеру его труды по механике (1736). Вскоре после этого вместе с Даниилом Бернулли и Маклореном он получил премию от Парижской академии наук за работу о морских приливах и отливах. К сожалению, напряженный научный труд ухудшил состояние его здоровья, что потребовало изменения климата. Поэтому Эйлер выехал в Берлин. Из года в год Эйлер печатал новые труды о движении планет и комет, о теории магнетизма, по баллистике. В 1748 году в Лозанне он издал свое главное произведение в трех томах «Введение в анализ бесконечно малых», в котором он собрал все свои прежние математические труды и статьи, написанные на протяжении многих лет. Эта работа укрепила позицию Эйлера как наиболее выдающегося математика. Почти все, что в настоящее время изучается по высшей алгебре и математическому анализу, включено в этот труд. Эйлер ввел в математику обозначение чисел « e » и « i », обобщенные координаты, удобные для описания вращательного движения, комплексные функции, комплексные переменные, в частности, известна формула Эйлера ($e^{i\varphi} = \cos \varphi + i \cdot \sin \varphi$), применение комплексных чисел при расчете интегралов, часто встречающиеся специальные функции гамма и бета, теорию дифференциальных уравнений. Эйлер является

также автором известной теоремы о том, что число ребер многогранника на 2 меньше числа его вершин и граней.

Знаменитую задачу о мостах в Кенигсберге тоже поставил Эйлер. Он занялся вопросом, можно ли пройти через 7 мостов, соединяющих районы города Кенигсберга с 2 островами на реке Преголе, так, чтобы не проходить по какому-либо из них два раза, проходя последовательно через все мосты, причем большой остров был соединен с каждым из берегов двумя мостами, а малый остров – одним мостом, оба острова были соединены между собой тоже одним мостом. Попытки, которые мы советуем сделать читателю, покажут, что решить эту задачу невозможно.

В 1766 году Эйлер возвратился в Россию. Екатерина II назначила ему постоянное жалование из собственных средств. «Я надеюсь, – сказала она, – что моя Академия возродится из пепла, когда к ней вернется великий человек». К сожалению, вскоре после приезда в Петербург Эйлер заболел и потерял второй глаз. Но его математический гений и великолепная память позволили ему продолжать работу. Формулы он писал мелом на доске, а своим друзьям диктовал новые работы. Таким образом, появились алгебра Эйлера, затем плод его тридцатилетнего труда – произведение на тему диоптрики и многие другие. Характерно, что гений Эйлера позволял ученому творить до поздней старости, о чем свидетельствует непрерывно растущее количество написанных им трудов. Эйлер написал свыше 800 работ, в том числе более половины по математике. Еще в день своей смерти он вел оживленный спор со своими сотрудниками. Эйлер умер 7 сентября 1783 года и похоронен на Смоленском кладбище в Петербурге, ныне его прах перенесен в Ленинградский Некрополь.

Один из сотрудников Эйлера говорил: «Математические формулы у Эйлера жили своей собственной жизнью и рассказывали ему важные и существенные данные о природе вещей. Ему было достаточно только коснуться их, как они из немых букв преображались в красноречивые фразы, дающие глубокий и значительный ответ на различные вопросы».

Великий французский математик Лаплас сказал о работах Эйлера: «Читайте, читайте Эйлера – он наш великий учитель».



КАРЛ ГУСТАВ ЯКОБ ЯКОБИ

(1804 – 1851)

Карл Густав Якоб Якоби родился в Потсдаме в семье берлинского банкира 10 декабря 1804 года. Его старший брат был известным и уважаемым физиком, профессором в Петербурге. Густав рано заинтересовался математикой. Много внимания любимой науке он посвятил в средней школе. После окончания школы он начал изучать математику в Берлине. Его научная карьера была стремительной. В возрасте всего лишь 21 года, осенью 1825 года, Якоби была присвоена ученая степень доктора. Весной следующего года он уехал в Кенигсберг и поступил на работу в университет.

По общему мнению, Якоби был прекрасным педагогом. Работая в университете, он углублял свои математические знания. В Кенигсберге в 1826 году ему было присвоено звание доцента, годом позже – звание экстраординарного профессора, в 1836 году он уже был членом Берлинской академии наук.

Якоби принадлежит числу создателей теории эллиптических функций. Он открыл ряд правил по теории чисел, линейной алгебре, исследовал дифференциальные уравнения динамики, занимался уравнениями первого порядка с частными производными. Работал над многими математическими проблемами. Трудно найти отрасль современной математики, в которой не работал бы Якоби. Основное его произведение «Fundamenta novae theoriae functionum ellipticarum». Несмотря на молодость, он был известным педагогом, которого высоко ценили.

В своей научной работе Якоби вновь занялся рядом проблем, оставленных Гауссом и Лагранжем. В частности, он закончил таблицы уравнения $X^n = A$, которые Гаусс разработал только для чисел до 89, расширив их на все числа до 1000. Якоби продолжил начатые Лагранжем работы по уравнению деления круга и работы Гаусса по теории чисел. Исследовал класс ортогональных многочленов, являющихся обобщением многочленов Лежандра.

В 1829 году Якоби начал работу в области теории аналитических функций. В 1834 году он напечатал свои труды, касающиеся квадратичных форм. Он интересовался также закономерностями во множестве целых чисел. Он высказал идею представления любого целого числа в качестве суммы квадратов целых чисел, пытался применить теорию эллиптических функций в аналитической геометрии, например, с помощью эллиптического интеграла, осуществил конформное преобразование эллипсоида в плоскость.

Весьма интенсивная научная работа быстро истощила молодой организм. В 1843 году Якоби оставил Кенигсберг и уехал лечиться в Италию. Оттуда ученый вернулся в Берлин, где работал еще несколько лет. Умер Якоби в Берлине 18 февраля 1851 года от оспы.

Полное собрание сочинений Якоби в 8 томах издано в 1881 – 1891 годах Берлинской академией наук.

Тем. план 1998, поз. 9

Учебное издание

Пирумов Ульянов Гайкович

ЧИСЛЕННЫЕ МЕТОДЫ

Редактор *Е.В. Лисовец*

ИБ № 317

Лицензия ЛР № 040211 от 7.04.97 г.
Подписано в печать 24.11.98.
Формат 60 × 84 1/16. Бумага газетная.
Гарнитура Таймс. Печать офсетная.
Усл.печ.л. 10,93. Уч.-изд.л. 12,88.
Заказ 2357. Тираж 1500 экз. С.75.

Издательство МАИ, 125871, Москва, Волоколамское шоссе, 4
Отпечатано
в типографии Издательства МАИ,
125871, Москва, Волоколамское шоссе, 4.